

BEST AVAILABLE COPY

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 December 2000 (28.12.2000)

PCT

(10) International Publication Number
WO 00/78954 A2

(51) International Patent Classification⁷: **C12N 15/12**,
C07K 14/47, 16/18, C12Q 1/68, G01N 33/68, A01K
67/027, A61K 38/17

[US/US]; 14244 Santiago Road, San Leandro, CA 94577
(US). **AZIMZAI, Yalda** [US/US]; 2045 Rock Springs
Drive, Hayward, CA 94545 (US). **TRAN, Bao** [US/US];
750 Salberg Avenue, Santa Clara, CA 95051 (US).

(21) International Application Number: PCT/US00/16766

(22) International Filing Date: 15 June 2000 (15.06.2000)

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/140,109 18 June 1999 (18.06.1999) US

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE,
ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP,
KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD,
SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ,
VN, YU, ZA, ZW.

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/140,109 (CIP)
Filed on 18 June 1999 (18.06.1999)

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **INCYTE
GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).

Published:

— *Without international search report and to be republished
upon receipt of that report.*

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **LAL, Preeti**
[IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US).
YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA
94087 (US). **TANG, Y., Tom** [CN/US]; 4230 Ranwick
Court, San Jose, CA 95118 (US). **BAUGHN, Mariah, R.**

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: HUMAN TRANSCRIPTIONAL REGULATOR PROTEINS

(57) Abstract: The invention provides human transcriptional regulator proteins (TXREG) and polynucleotides which identify and encode TXREG. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of TXREG.

WO 00/78954 A2

HUMAN TRANSCRIPTIONAL REGULATOR PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human transcriptional
5 regulator proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell
proliferative, autoimmune/inflammatory, and developmental disorders.

BACKGROUND OF THE INVENTION

Multicellular organisms are comprised of diverse cell types that differ dramatically both in
10 structure and function. The identity of a cell is determined by its characteristic pattern of gene
expression, and different cell types express overlapping but distinctive sets of genes throughout
development. Spatial and temporal regulation of gene expression is critical for the control of cell
proliferation, cell differentiation, apoptosis, and other processes that contribute to organismal
development. Furthermore, gene expression is regulated in response to extracellular signals that
15 mediate cell-cell communication and coordinate the activities of different cell types. Appropriate
gene regulation also ensures that cells function efficiently by expressing only those genes whose
functions are required at a given time.

Transcription Factors

Transcriptional regulatory proteins are essential for the control of gene expression. Some of
20 these proteins function as transcription factors that initiate, activate, repress, or terminate gene
transcription. Transcription factors generally bind to the promoter, enhancer, and upstream
regulatory regions of a gene in a sequence-specific manner, although some factors bind regulatory
elements within or downstream of a gene's coding region. Transcription factors may bind to a
specific region of DNA singly or as a complex with other accessory factors. (Reviewed in Lewin, B.
25 (1990) Genes IV, Oxford University Press, New York, NY, and Cell Press, Cambridge, MA, pp. 554-
570.)

The double helix structure and repeated sequences of DNA create topological and chemical
features which can be recognized by transcription factors. These features are hydrogen bond donor
and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches
30 of sequence which induce distinct bends in the helix. Typically, transcription factors recognize
specific DNA sequence motifs of about 20 nucleotides in length. Multiple, adjacent transcription
factor-binding motifs may be required for gene regulation.

Many transcription factors incorporate DNA-binding structural motifs which comprise either
 α helices or β sheets that bind to the major groove of DNA. Four well-characterized structural motifs
35 are the helix-turn-helix, zinc finger, leucine zipper, and helix-loop-helix. Proteins containing these

motifs may act alone as monomers, or they may form homo- or heterodimers that interact with DNA.

The helix-turn-helix motif consists of two α helices connected at a fixed angle by a short chain of amino acids. One of the helices binds to the major groove. Helix-turn-helix motifs are exemplified by the homeobox motif which is present in homeodomain proteins. These proteins are critical for specifying the anterior-posterior body axis during development and are conserved throughout the animal kingdom. The Antennapedia and Ultrabithorax proteins of Drosophila melanogaster are prototypical homeodomain proteins (Pabo, C.O. and R.T. Sauer (1992) Ann. Rev. Biochem. 61:1053-1095).

The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern include the C2H2-type, C4-type, and C3HC4-type ("RING" finger) zinc fingers, and the PHD domain (Lewin, supra ; Aasland, R. et al. (1995) Trends Biochem. Sci 20:56 - 59). Zinc finger proteins each contain an α helix and an antiparallel β sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the α helix and by the second, third, and sixth residues of the α helix. The zinc finger motif may be repeated in a tandem array within a protein, such that the α helix of each zinc finger in the protein makes contact with the major groove of the DNA double helix. This repeated contact between the protein and the DNA produces a strong and specific DNA-protein interaction. The strength and specificity of the interaction can be regulated by the number of zinc finger motifs within the protein.

The leucine zipper motif comprises a stretch of amino acids rich in leucine which can form an amphipathic α helix. This structure provides the basis for dimerization of two leucine zipper proteins. The region adjacent to the leucine zipper is usually basic, and upon protein dimerization, is optimally positioned for binding to the major groove. Proteins containing such motifs are generally referred to as bZIP transcription factors. The leucine zipper motif is found in the proto-oncogenes Fos and Jun, which comprise the heterodimeric transcription factor AP1, involved in cell growth and the determination of cell lineage (Papavassiliou, A. G. (1995) N. Engl. J. Med. 332:45-47).

The helix-loop-helix motif (HLH) consists of a short α helix connected by a loop to a longer α helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The oncogene Myc, a transcription factor that activates genes required for cellular proliferation, contains a prototypical HLH motif.

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized (Faisst, S. and S. Meyer (1992) Nucl. Acids Res. 20:3-26). These include the forkhead motif, found in transcription factors involved in development and oncogenesis (Hacker, U. et al. (1995) EMBO J. 14:5306-5317).

Chromatin Associated Proteins

In the nucleus, DNA is packaged into chromatin, the compact organization of which limits the accessibility of DNA to transcription factors and plays a key role in gene regulation (Lewin, supra, pp. 409-410). The compact structure of chromatin is determined and influenced by chromatin-associated proteins such as the histones, the high mobility group (HMG) proteins, helicases, and the chromodomain proteins. There are five classes of histones, H1, H2A, H2B, H3, and H4, all of which are highly basic, low molecular weight proteins. The fundamental unit of chromatin, the nucleosome, consists of 200 base pairs of DNA associated with two copies each of H2A, H2B, H3, and H4. H1 links adjacent nucleosomes. HMG proteins are low molecular weight, non-histone proteins that may play a role in unwinding DNA and stabilizing single-stranded DNA. Helicases, which are DNA-dependent ATPases, unwind DNA, allowing access for transcription factors. Chromodomain proteins play a key role in the formation of highly compacted heterochromatin, which is transcriptionally silent.

Diseases and disorders related to gene regulation

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) *Cancer Surv.* 15:89-104). The zinc finger-type transcriptional regulator WT1 is a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, supra). Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of a transcriptional regulator with the regulatory regions of a second unrelated gene. In Burkitt's lymphoma, for example, the transcription factor Myc is translocated to the immunoglobulin heavy chain locus, greatly enhancing Myc expression and resulting in rapid cell growth leading to leukemia (Latchman, D. S. (1996) *N. Engl. J. Med.* 334:28-33).

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections (Isselbacher et al. Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996). The causative gene for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) was recently isolated and found to encode a protein with two PHD-type zinc finger motifs (Bjorses, P. et al. (1998) *Hum. Mol. Genet.* 7:1547-1553).

Furthermore, the generation of multicellular organisms is based upon the induction and

coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development can result in developmental disorders. Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3), and Townes-Brocks syndrome, characterized by anal, renal, limb, and ear abnormalities (SALL1) (Engelkamp, D. and V. van Heyningen (1996) *Curr. Opin. Genet. Dev.* 6:334-342; Kohlhase, J. et al. (1999) *Am. J. Hum. Genet.* 64:435-445).

The discovery of new human transcriptional regulator proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, and developmental disorders.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, human transcriptional regulator proteins, referred to collectively as "TXREG" and individually as "TXREG-1," "TXREG-2," "TXREG-3," "TXREG-4," "TXREG-5," "TXREG-6," "TXREG-7," "TXREG-8," "TXREG-9," "TXREG-10," "TXREG-11," "TXREG-12," "TXREG-13," "TXREG-14," "TXREG-15," "TXREG-16," "TXREG-17," "TXREG-18," "TXREG-19," "TXREG-20," "TXREG-21," "TXREG-22," "TXREG-23," "TXREG-24," "TXREG-25," "TXREG-26," "TXREG-27," "TXREG-28," "TXREG-29," "TXREG-30," "TXREG-31," and "TXREG-32." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-32.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected

from the group consisting of SEQ ID NO:1-32. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-32. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:33-64.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter
5 sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of
10 SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid
15 sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the
20 group consisting of SEQ ID NO:1-32. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a
25 polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino
30 acid sequence selected from the group consisting of SEQ ID NO:1-32.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID
35 NO:33-64, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence

complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide
5 sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the
10 sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the
15 probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, b) a naturally occurring polynucleotide sequence having at least
20 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and,
25 optionally, if present, the amount thereof.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and a pharmaceutically acceptable excipient. In one embodiment, the pharmaceutical composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The invention additionally
35 provides a method of treating a disease or condition associated with decreased expression of

functional TXREG, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TXREG, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TXREG, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a)

combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:33-64, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding TXREG.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of TXREG.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding TXREG were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and

polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

5 Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

10 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

15 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing
20 the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TXREG" refers to the amino acid sequences of substantially purified TXREG obtained from
25 any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TXREG. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TXREG either by directly interacting with
30 TXREG or by acting on components of the biological pathway in which TXREG participates.

An "allelic variant" is an alternative form of the gene encoding TXREG. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to
35 allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides.

Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding TXREG include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TXREG or a polypeptide with at least one functional characteristic of TXREG. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TXREG, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TXREG. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TXREG. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TXREG is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of TXREG. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TXREG either by directly interacting with TXREG or by acting on components of the biological pathway in which TXREG participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TXREG polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or

oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize
5 the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures
10 on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as
15 phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring
20 nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"
25 refers to the capability of the natural, recombinant, or synthetic TXREG, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,
30 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TXREG or fragments of TXREG may
35 be employed as hybridization probes. The probes may be stored in freeze-dried form and may be

associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
20	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
25	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
30	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
35	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule.

5 A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

10 A "fragment" is a unique portion of TXREG or the polynucleotide encoding TXREG which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10,
15 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the
20 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:33-64 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:33-64, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:33-64 is useful, for
25 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:33-64 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:33-64 and the region of SEQ ID NO:33-64 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-32 is encoded by a fragment of SEQ ID NO:33-64. A fragment
30 of SEQ ID NO:1-32 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-32. For example, a fragment of SEQ ID NO:1-32 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-32. The precise length of a fragment of SEQ ID NO:1-32 and the region of SEQ ID NO:1-32 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended
35 purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

5 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous
10 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes
15 in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some
20 alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e
25 sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

30 Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

35 *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

5 Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment
10 length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

 “Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome replication, segregation and maintenance.

15 The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

 “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific
20 hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive
25 conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

30 Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and
35 conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al.,

1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TXREG which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TXREG which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TXREG. For example, modulation

may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TXREG.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TXREG may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TXREG.

"Probe" refers to nucleic acid sequences encoding TXREG, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for

example Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs
5 can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to
10 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from
15 megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection
20 programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both
25 unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

30 A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have
35 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a

recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding TXREG, or fragments thereof, or TXREG itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

5 A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

15 A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

20 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may

35

have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human transcriptional regulator proteins (TXREG), the polynucleotides encoding TXREG, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative, autoimmune/inflammatory, and developmental disorders.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding TXREG. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each TXREG were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The Incyte clones in column 5 were used to assemble the consensus nucleotide sequence of each TXREG and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The

methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding TXREG. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:33-64 and to distinguish between SEQ ID NO:33-64 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express TXREG as a fraction of total tissues expressing TXREG. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing TXREG as a fraction of total tissues expressing TXREG. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding TXREG were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

SEQ ID NO:33 maps to chromosome 1 within the interval from 199.2 to 203.0 centiMorgans, to chromosome 6 within the interval from 59.6 to 73.9 centiMorgans, and to chromosome 13 within the interval from 112.8 to 117.5 centiMorgans. The interval on chromosome 6 from 59.6 to 73.9 centiMorgans also contains genes associated with methylmalonic CoA mutase deficiency and retinal degeneration. The interval on chromosome 13 from 112.8 to 117.5 centiMorgans also contains genes associated with Oguchi disease (night blindness) and Factor X deficiency. SEQ ID NO:34 maps to chromosome 13 within the interval from 112.8 to 117.5 centiMorgans. This interval also contains genes associated with Oguchi disease (night blindness) and Factor X deficiency. SEQ ID NO:35 maps to chromosome 12 within the interval from 113.3 to 126.1 centiMorgans. This interval also contains genes associated with spinocerebellar ataxia, mevalonate kinase deficiency, alcohol intolerance, and myocardial hypertrophy. SEQ ID NO:36 maps to chromosome 1 within the interval from 155.2 to 157.4 centiMorgans, and to chromosome 16 within the interval from 83.7 to 86.6 centiMorgans. The interval on chromosome 1 from 155.2 to 157.4 centiMorgans also contains genes associated with leukemia and adrenal hyperplasia. The interval on chromosome 16 from 83.7 to 86.6 centiMorgans also contains a gene associated with cortisol 11-beta-keto reductase deficiency. SEQ ID NO:38 maps to chromosome 9 within the interval from 59.9 to 64.5 centiMorgans. SEQ ID NO:40 maps to chromosome 18 within the interval from 61.2 to 63.2 centiMorgans. SEQ ID NO:44 maps to chromosome 2 within the interval from 180.6 to 188.2 centiMorgans. This interval also contains a gene associated with glutamate decarboxylase deficiency. SEQ ID NO:45 maps to

chromosome 13 within the interval from 112.8 to 117.5 centiMorgans. This interval also contains genes associated with Oguchi disease (night blindness) and Factor X deficiency. SEQ ID NO:47 maps to chromosome 8 within the interval from 75.0 to 90.2 centiMorgans. This interval also contains genes associated with branchiootorenal dysplasia and Zellweger syndrome. SEQ ID NO:61 maps to chromosome 5 within the interval from 63.9 to 69.6 centiMorgans. SEQ ID NO:62 maps to chromosome 7 within the interval from 120.7 to 123.9 centiMorgans. This interval also contains genes associated with lipoamide dehydrogenase deficiency, neonatal cutis laxa, and tumor suppression. SEQ ID NO:64 maps to chromosome 1 within the interval from 157.4 to 186.4 centiMorgans, to chromosome 5 within the interval from 175.3 to 182.4 centiMorgans, and to chromosome 14 within the interval from 7.5 to 21.9 centiMorgans. The interval on chromosome 1 from 157.4 to 186.4 centiMorgans also contains genes associated with autoimmune diseases, leukemia, and Gaucher disease. The interval on chromosome 14 from 7.5 to 21.9 centiMorgans also contains genes associated with apoptosis, hypertrophic cardiomyopathy, and oculopharyngeal muscular dystrophy.

15 The invention also encompasses TXREG variants. A preferred TXREG variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TXREG amino acid sequence, and which contains at least one functional or structural characteristic of TXREG.

20 The invention also encompasses polynucleotides which encode TXREG. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:33-64, which encodes TXREG. The polynucleotide sequences of SEQ ID NO:33-64, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

25 The invention also encompasses a variant of a polynucleotide sequence encoding TXREG. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TXREG. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:33-64 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:33-64. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TXREG.

35 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TXREG, some bearing minimal

similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the
5 polynucleotide sequence of naturally occurring TXREG, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TXREG and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TXREG under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding
10 TXREG or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TXREG and its derivatives without altering the encoded amino acid
15 sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TXREG and TXREG derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell
20 systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TXREG or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:33-64 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and
25 S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment
30 of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer
35 system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI

CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g.,

5 Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TXREG may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,

10 such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising

15 a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region

20 of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed

25 using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been

30 size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze

35 the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary

sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire
5 process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TXREG may be cloned in recombinant DNA molecules that direct expression of
10 TXREG, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TXREG.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TXREG-encoding sequences for a variety of purposes including, but
15 not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

20 The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TXREG, such as its biological or enzymatic activity or its ability
25 to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial"
30 breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable
35 manner.

In another embodiment, sequences encoding TXREG may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.)

Alternatively, TXREG itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of TXREG, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TXREG, the nucleotide sequences encoding TXREG or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TXREG. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TXREG. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TXREG and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TXREG and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A

Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TXREG. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TXREG. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TXREG can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TXREG into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TXREG are needed, e.g. for the production of antibodies, vectors which direct high level expression of TXREG may be used. For example, vectors

containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TXREG. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such
5 vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, supra; and Scorer, supra.)

Plant systems may also be used for expression of TXREG. Transcription of sequences encoding TXREG may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used
10 alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw
15 Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TXREG may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain
20 infective virus which expresses TXREG in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of
25 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression
30 of TXREG in cell lines is preferred. For example, sequences encoding TXREG can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to
35 confer resistance to a selective agent, and its presence allows growth and recovery of cells which

successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk⁻* and *apr⁻* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TXREG is inserted within a marker gene sequence, transformed cells containing sequences encoding TXREG can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TXREG under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TXREG and that express TXREG may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TXREG using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TXREG is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,

e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

5 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TXREG include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TXREG, or any fragments thereof, may be cloned into a vector
10 for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for
15 ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

 Host cells transformed with nucleotide sequences encoding TXREG may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence
20 and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TXREG may be designed to contain signal sequences which direct secretion of TXREG through a prokaryotic or eukaryotic cell membrane.

 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of
25 the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the
30 American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TXREG may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TXREG protein
35 containing a heterologous moiety that can be recognized by a commercially available antibody may

facilitate the screening of peptide libraries for inhibitors of TXREG activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TXREG encoding sequence and the heterologous protein sequence, so that TXREG may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TXREG may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

TXREG of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TXREG. At least one and up to a plurality of test compounds may be screened for specific binding to TXREG. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TXREG, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TXREG binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TXREG, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TXREG or cell membrane fractions which contain TXREG are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TXREG or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TXREG, either in solution or affixed to a solid support, and detecting the binding of TXREG to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a
5 labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TXREG of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TXREG. Such compounds may include agonists, antagonists, or partial
10 or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TXREG activity, wherein TXREG is combined with at least one test compound, and the activity of TXREG in the presence of a test compound is compared with the activity of TXREG in the absence of the test compound. A change in the activity of TXREG in the presence of the test compound is indicative of a compound that modulates the activity of TXREG. Alternatively, a test compound is
15 combined with an in vitro or cell-free system comprising TXREG under conditions suitable for TXREG activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TXREG may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TXREG or their mammalian homologs
20 may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of
25 interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids
30 Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

35 Polynucleotides encoding TXREG may also be manipulated in vitro in ES cells derived from

human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

- 5 Polynucleotides encoding TXREG can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TXREG is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and
10 treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TXREG, e.g., by secreting TXREG in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

- 15 Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TXREG and human transcriptional regulator proteins. In addition, the expression of TXREG is closely associated with cell proliferation and inflammation. Therefore, TXREG appears to play a role in cell proliferative, autoimmune/inflammatory, and developmental disorders. In the treatment of disorders associated with increased TXREG expression or activity, it is desirable
20 to decrease the expression or activity of TXREG. In the treatment of disorders associated with decreased TXREG expression or activity, it is desirable to increase the expression or activity of TXREG.

- Therefore, in one embodiment, TXREG or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or
25 activity of TXREG. Examples of such disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of
30 the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis,
35 autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-

candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss.

In another embodiment, a vector capable of expressing TXREG or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TXREG including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified TXREG in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TXREG including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TXREG may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TXREG including, but not limited to, those listed above.

In a further embodiment, an antagonist of TXREG may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TXREG. Examples of such disorders include, but are not limited to, those cell proliferative, autoimmune/inflammatory, and developmental disorders described above. In one aspect, an antibody which specifically binds TXREG may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TXREG.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TXREG may be administered to a subject to treat or prevent a disorder associated with

increased expression or activity of TXREG including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made
5 by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TXREG may be produced using methods which are generally known in the
10 art. In particular, purified TXREG may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TXREG. Antibodies to TXREG may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit
15 dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TXREG or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral
20 gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TXREG have an amino acid sequence consisting of at least about 5 amino acids, and generally will
25 consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TXREG amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TXREG may be prepared using any technique which provides for
30 the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

35 In addition, techniques developed for the production of "chimeric antibodies," such as the

splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TXREG-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TXREG may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TXREG and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TXREG epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TXREG. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of TXREG-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TXREG epitopes, represents the average affinity, or avidity, of the antibodies for TXREG. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TXREG epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TXREG-antibody complex must withstand rigorous manipulations. Low-affinity antibody

preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TXREG, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TXREG-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding TXREG, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TXREG. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TXREG. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding TXREG may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-

linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and Somia, N. (1997) Nature 389:239-242), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TXREG expression or regulation causes disease, the expression of TXREG from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TXREG are treated by constructing mammalian expression vectors encoding TXREG and introducing these vectors by mechanical means into TXREG-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TXREG include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TXREG may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. U.S.A. 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V.

and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TXREG from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver
5 polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

10 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TXREG expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TXREG under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences
15 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.
20 (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant")
discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by
25 reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-
30 2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TXREG to cells which have one or more genetic abnormalities with respect to the expression of TXREG. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to
35 be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas

(Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544; and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both
5 incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TXREG to target cells which have one or more genetic abnormalities with respect to the expression of TXREG. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TXREG to cells of the central nervous system, for which HSV has
10 a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is
15 hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al.
20 (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

25 In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TXREG to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotech.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid
30 proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for TXREG into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TXREG-coding RNAs and the synthesis of high levels of TXREG in vector transduced cells. While
35 alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a

persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TXREG into a variety of cell types. The specific transduction of a subset of
5 cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions
10 -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E.
15 and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme
20 molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TXREG.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA,
25 GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

30 Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TXREG. Such DNA sequences may be incorporated into a wide variety of
35 vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA

constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TXREG. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TXREG expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TXREG may be therapeutically useful, and in the treatment of disorders associated with decreased TXREG expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TXREG may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TXREG is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TXREG are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TXREG. The amount of hybridization may be quantified, thus

forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific
5 polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide
10 nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells
15 taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of
20 such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and
25 proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such pharmaceutical compositions may consist of TXREG, antibodies to TXREG, and mimetics, agonists, antagonists, or inhibitors of TXREG.

The pharmaceutical compositions utilized in this invention may be administered by any
30 number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Pharmaceutical compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the
35 patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol

delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of
5 administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

10 Specialized forms of pharmaceutical compositions may be prepared for direct intracellular delivery of macromolecules comprising TXREG or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TXREG or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been
15 found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration
20 range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TXREG or fragments thereof, antibodies of TXREG, and agonists, antagonists or inhibitors of TXREG, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be
25 determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and
30 animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the
35 subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the

active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art.

- Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

- In another embodiment, antibodies which specifically bind TXREG may be used for the diagnosis of disorders characterized by expression of TXREG, or in assays to monitor patients being treated with TXREG or agonists, antagonists, or inhibitors of TXREG. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TXREG include methods which utilize the antibody and a label to detect TXREG in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

- A variety of protocols for measuring TXREG, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TXREG expression. Normal or standard values for TXREG expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to TXREG under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TXREG expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

- In another embodiment of the invention, the polynucleotides encoding TXREG may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TXREG may be correlated

with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TXREG, and to monitor regulation of TXREG levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TXREG or closely related molecules may be used to identify nucleic acid sequences which encode TXREG. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TXREG, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TXREG encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:33-64 or from genomic sequences including promoters, enhancers, and introns of the TXREG gene.

Means for producing specific hybridization probes for DNAs encoding TXREG include the cloning of polynucleotide sequences encoding TXREG or TXREG derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TXREG may be used for the diagnosis of disorders associated with expression of TXREG. Examples of such disorders include, but are not limited to, a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis,

erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss. The polynucleotide sequences encoding TXREG may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TXREG expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TXREG may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TXREG may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TXREG in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TXREG, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TXREG, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified

polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,
5 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or
10 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

15 Additional diagnostic uses for oligonucleotides designed from the sequences encoding TXREG may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TXREG, or a fragment of a polynucleotide complementary to the polynucleotide encoding TXREG, and will be employed under optimized conditions for identification of a specific gene or
20 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TXREG may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic
25 disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TXREG are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause
30 differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the
35 sequence of individual overlapping DNA fragments which assemble into a common consensus

sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

5 Methods which may also be used to quantify the expression of TXREG include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of
10 interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large
15 numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and
20 monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

25 In another embodiment, antibodies specific for TXREG, or TXREG or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

 Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci.
30 USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

35 In another embodiment of the invention, nucleic acid sequences encoding TXREG may be

used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during

5 chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J.

10 (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic

15 map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TXREG on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

20 In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery

25 techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to

30 translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TXREG, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

35 between TXREG and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TXREG, or fragments thereof, and washed. Bound TXREG is then detected by methods well known in the art. Purified TXREG can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TXREG specifically compete with a test compound for binding TXREG. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TXREG.

In additional embodiments, the nucleotide sequences which encode TXREG may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No.60/140,109 are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A⁺) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was

isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP
5 vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-
10 1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUEScript plasmid (Stratagene), PSORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant
15 plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using
20 at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

25 Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence
30 scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ
35 Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB

2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VI.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS,

DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:33-64. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding TXREG occurred. Analysis involved the categorization of cDNA libraries by

organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Chromosomal Mapping of TXREG Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:33-64 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:33-64 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO., to that map location.

The genetic map locations of SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:47, SEQ ID NO:61, SEQ ID NO:62, and SEQ ID NO:64 are described in The Invention as ranges, or intervals, of human chromosomes. More than one map location is reported for SEQ ID NO:33, SEQ ID NO:36, and SEQ ID NO:64, indicating that previously mapped sequences having similarity, but not complete identity, to SEQ ID NO:33, SEQ ID NO:36, and SEQ ID NO:64 were assembled into their respective clusters. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Diseases associated with the public and Incyte sequences located within the indicated intervals are also reported in the Invention where applicable.

VI. Extension of TXREG Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:33-64 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other

primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and β -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

10 In like manner, the polynucleotide sequences of SEQ ID NO:33-64 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:33-64 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

25 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999),

supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37 °C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85 °C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5
5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR
10 Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average
15 concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate
20 buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60 °C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample
25 mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash
30 buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines
35 at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is

focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

5 In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is
10 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that
15 location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

20 The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and
25 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The
30 software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

IX. Complementary Polynucleotides

Sequences complementary to the TXREG-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TXREG. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same
35 procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are

designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TXREG. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TXREG-encoding transcript.

X. Expression of TXREG

Expression and purification of TXREG is achieved using bacterial or virus-based expression systems. For expression of TXREG in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TXREG upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TXREG in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TXREG by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TXREG is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TXREG at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TXREG obtained by these methods can be used directly in the assays shown in Examples XI and XV.

XI. Demonstration of TXREG Activity

TXREG activity is measured by its ability to stimulate transcription of a reporter gene (Liu, H.Y. et al. (1997) EMBO J. 16(17):5289-5298). The assay employs a well characterized reporter gene construct, LexA_{op}-LacZ, that consists of LexA DNA transcriptional control elements (LexA_{op}) fused to sequences encoding the *E. coli* LacZ enzyme. The methods for constructing and expressing fusions genes, introducing them into cells, and measuring LacZ enzyme activity, are well known to those skilled in the art. Sequences encoding TXREG are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-TXREG, consisting of TXREG and a DNA binding domain derived from the LexA transcription factor. The resulting plasmid, encoding a LexA-TXREG fusion protein, is introduced into yeast cells along with a plasmid containing the LexA_{op}-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-TXREG transfected cells, relative to control cells, is proportional to the amount of transcription stimulated by the TXREG.

XII. Functional Assays

TXREG function is assessed by expressing the sequences encoding TXREG at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) *Flow Cytometry*, Oxford, New York NY.

The influence of TXREG on gene expression can be assessed using highly purified

populations of cells transfected with sequences encoding TXREG and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake
5 Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TXREG and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIII. Production of TXREG Specific Antibodies

TXREG substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,
10 Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TXREG amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for
15 selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase
20 immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TXREG activity by, for example, binding the peptide or TXREG to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIV. Purification of Naturally Occurring TXREG Using Specific Antibodies

25 Naturally occurring or recombinant TXREG is substantially purified by immunoaffinity chromatography using antibodies specific for TXREG. An immunoaffinity column is constructed by covalently coupling anti-TXREG antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

30 Media containing TXREG are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TXREG (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TXREG binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TXREG is collected.

XV. Identification of Molecules Which Interact with TXREG

35

TXREG, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TXREG, washed, and any wells with labeled TXREG complex are assayed. Data obtained using different

5 concentrations of TXREG are used to calculate values for the number, affinity, and association of TXREG with the candidate molecules.

Alternatively, molecules interacting with TXREG are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, *Nature* 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

10 TXREG may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

15. Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention

20 which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	33	091502	HYPONOB01	091502H1 (HYPONOB01), 427507R6 (BLADNOT01), 1621915F6 (BRAITUT13), 1971979H1 (UCMCL5T01), 2114854H1 (BRAITUT03), 2187734F6 (PROSNOT26), 2189475F6 (PROSNOT26), 2232396F6 (PROSNOT16), 2516773F6 (LIVRTUT04), 2599557F6 (UTRSNOT10), 4347694H1 (TLYMTXT01), 4821742H1 (PROSTUT17), 4888937H1 (PROSTMT05), 5731093H1 (KIDCTMT01)
2	34	763816	LUNGNOT04	715015F1 (PROSTUT01), 763816H1 (LUNGNOT04), 1910519F6 (CONNTUT01), 2583937F6 (BRAITUT22), 2879360H1 (UTRSTUT05), 3040246H1 (BRSTNOT16), 3581025F6 (293TF3T01), 3581025T6 (293TF3T01), 4181312H1 (SINITUT03), 5098614H2 (EPIMNON05)
3	35	961184	BRSTTUT03	155935R6 (THP1PLB02), 614996R6 (COLNTUT02), 961184H1 (BRSTTUT03), 961184R2 (BRSTTUT03), 1256271F6 (MENITUT03), 1470756F6 (PANCUTUT02), 1680384H1 (STOMFET01), 2419005F6 (HNT3AZT01), 2838738F6 (DRGLNOT01), 4625178H1 (FIBRTXT02)
4	36	1255525	MENITUT03	149604R6 (FIBRNGT02), 996734R1 (KIDNTUT01), 999068T6 (KIDNTUT01), 1761029R6 (PITUNOT03), 2054882R6 (BEPINOT01), 2238879F6 (PANCUTUT02), 2985829H1 (CARGDIT01), 4827019H1 (BLADDIT01)
5	37	1297447	BRSTNOT07	1297447F6 (BRSTNOT07), 1297447H1 (BRSTNOT07), 1461457R1 (PANCNOT04), 1944275T6 (PITUNOT01), 4636727H1 (MYEPTXT01)
6	38	1441094	THYRNOT03	536572T6 (LNODNOT02), 879951R1 (THYRNOT02), 1441094H1 (THYRNOT03), 1441094T6 (THYRNOT03), 1755891F6 (LIVRTUT01), 2808466F6 (BLADTUT08), 3125460F6 (LNODNOT05), 3297532F6 (TLYJINT01), 4200612H1 (BRAITUT29), 4727419H1 (GBLADIT01)
7	39	1479382	CORPNOT02	1330922F1 (PANCNOT07), 1479382F1 (CORPNOT02), 1479382H1 (CORPNOT02), 1483066F6 (CORPNOT02), 2900979F6 (DRGCNOT01), 3041118F7 (BRSTNOT16), 3779139H1 (BRSTNOT27), 4171269H1 (SINTNOT21), 5174156H1 (EPIBTXT01)
8	40	1503131	BRAITUT07	1503131F1 (BRAITUT07), 1503131H1 (BRAITUT07), 1691039F6 (PROSTUT10), 2051545T6 (LIVRFET02), 3393930H1 (LUNGNOT28), 4786813H1 (BRATNOT03), SBKA01094F1

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
9	41	1594803	BRAINOT14	549607R6 (BEPINOT01), 984580R1 (LVENNOT03), 1300947F1 (BRSTNOT07), 1479450F1 (CORPNOT02), 1594803H1 (BRAINOT14), 1865083F6 (PROSNOT19), 2210749F6 (SINTFET03), 2517123F7 (LIVRTUT04), 4938329H1 (EPIMNON04)
10	42	1736129	COLNNOT22	935470R1 (CERVNOT01), 1673992F6 (BLADNOT05), 1736129H1 (COLNNOT22), 1807715T6 (SINTNOT13), 3074440H1 (BONEUNT01)
11	43	1874312	LEUKNOT02	1874312F6 (LEUKNOT02), 1874312H1 (LEUKNOT02)
12	44	1969301	BRSTNOT04	775854R1 (COLNNOT05), 775974T1 (COLNNOT05)
13	45	1986873	LUNGAST01	715015F1 (PROSTUT01), 980391H1 (TONGTUT01), 1986873H1 (LUNGAST01), 2655130H1 (THYMNOT04), 2879360H1 (UTRSTUT05), 3134676H1 (SMCCNOT01), 3581025T6 (293TF3T01), 4181312H1 (SINITUT03), 4376343H1 (CONFNOT03), 5098614H2 (EPIMNON05), 5154716H1 (HEARFET03)
14	46	2010820	TESTNOT03	041893R6 (TBLYNOT01), 1866348F6 (THP1NOT01), 2010820H1 (TESTNOT03), 2496529H1 (ADRETUT05), 2747406F6 (LUNGUTUT11), 3393961F6 (LUNGNOT28), 3393961T6 (LUNGNOT28), 3524649H1 (ESOGTUN01), 3805523H1 (BLADTUT03)
15	47	2013818	TESTNOT03	415777R6 (BRSTNOT01), 901350R1 (BRSTTUT03), 1401608F6 (BRAITUT08), 1682956F7 (PROSNOT15), 1978792R6 (LUNGUTUT03), 2013818H1 (TESTNOT03), 3815325H1 (TONSNOT03), 4418093H1 (LIVRDIT02)
16	48	2302032	BRSTNOT05	1239431R6 (LUNGUTUT02), 2302032H1 (BRSTNOT05), 3281201F6 (STOMFET02)
17	49	2326109	OVARNOT02	1448156F6 (PLACNOT02), 1448156R1 (PLACNOT02), 2326109H1 (OVARNOT02), 2739124H1 (OVARNOT09), 2906494F6 (THYMNOT05), 4029574H1 (BRAINOT23), SCCA01806V1, SCCA02478V1, SCCA01171V1, SCCA05410V1, SCCA05970V1

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
18	50	2354751	LUNGNOT20	1510241F1 (LUNGNOT14), 1510241T1 (LUNGNOT14), 1708367F6 (PROSNOT16), 2354751F6 (LUNGNOT20), 2354751H1 (LUNGNOT20), 2630810H1 (COLNTUT15), 2691746F6 (LUNGNOT23), 2724416F6 (LUNGNOT10), 3043817H1 (HEAANOT01), 3089441F6 (HEAANOT03), 3089441T6 (HEAANOT03), 3228022H1 (COTRNOT01), 3428901H1 (SKINNOT04), 3687519F6 (HEAANOT01), 3697528H1 (SININOT05), 4898377H1 (OVARDT01), 5169814H1 (MUSCDMT01)
19	51	2378058	ISLTNOT01	276403R6 (TESTNOT03), 2378058H1 (ISLTNOT01), 2378058T6 (ISLTNOT01), 5285702H1 (TESTNON04)
20	52	2595747	OVRTUT02	077848R1 (SYNORAB01), 541248T6 (LNODNOT02), 1481520F6 (CORPNOT02), 2110927H1 (BRAITUT03), 2555938F6 (THYMNOT03), 2595747H1 (OVRTUT02), 3508070H1 (CONCNOT01)
21	53	2634391	COLNTUT15	342418T7 (NEUTFMT01), 991294R6 (COLNNOT11), 1997874X17F1 (BRSTTUT03), 2634391H1 (COLNTUT15), 3819551H1 (BONSTUT01)
22	54	2637522	BONTNOT01	2637522H1 (BONTNOT01), 3028034F6 (HEARFET02), 3042665H1 (HEAANOT01), 3573661F6 (BRONNOT01)
23	55	2650980	LUNGUT12	1848956T6.comp (LUNGFET03), 2345947H1 (TESTTUT02), 2396384F6 (THPLAZT01), 2396384T6.comp (THPLAZT01), 2515666X15C1 (LIVRTUT04), 2650980F6 (LUNGUT12), 2650980H1 (LUNGUT12), 2831227F6 (TLYMNOT03), 3861763H1 (LNODNOT03)
24	56	2939607	THYMFET02	453068X72 (TLYMNOT02), 2939607H1 (THYMFET02), 2939607T6 (THYMFET02), SBZA02575V1, SBZA05411V1
25	57	3098421	CERVNOT03	698595R6 (SYNORAT03), 959844R6 (BRSTTUT03), 3098421H1 (CERVNOT03), 3098421T6 (CERVNOT03), 3199392T6 (PENCNOT02), 3220706H1 (COLNNON03), 4583948H1 (OVARNOT13)
26	58	3296650	TLYJINT01	3296650H1 (TLYJINT01), SCCA00414V1, SCCA01718V1, SCCA04789V1, SCCA02483V1
27	59	3687719	HEAANOT01	2581531T6.comp (KIDNTUT13), 2593470F6 (OVRTUT02), 2799510F6 (PENCNOT01), 3448159H1 (UTRSNON03), 3687719F6 (HEAANOT01), 3687719H1 (HEAANOT01)

Table 1 (cont.)

Protein SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
28	60	3774188	BRSTNOT25	1397514F6 (BRAITUT08), 1905085F6 (OVARNOT07), 3557288H1 (LUNGNOT31), 3657636H1 (ENDPNOT02), 3774188H1 (BRSTNOT25)
29	61	4349106	TLYMTXT01	700322X11 (SYNORAT03), 789119H1 (PROSTUT03), 2117586T6 (BRSTTUT02), 3349655H1 (BRAITUT24), 3733533H1 (SMCCNOS01), 4349106H1 (TLYMTXT01)
30	62	4834217	BRAVTXT03	237019R6 (SINTNOT02), 943694T1 (ADREN0T03), 997161R2 (KIDNTUT01), 1393882F1 (THYRN0T03), 1506402F1 (BRAITUT07), 1843218R6 (COLNNOT08), 3095559H1 (CERVNOT03), 3360677F6 (PROSTUT16)
31	63	5156094	BRSTTMT02	592855H1 (BRAVUNT02), 791914F1 (PROSTUT03), 2870955T6 (THYRN0T10), 3209867T6 (BLADNOT08), 5156094H1 (BRSTTMT02), SBFA04503F1, SBFA01606F1, SEFA03681F1, SBFA00213F1
32	64	5665139	BRAUNOT01	063469F1 (PLACNOB01), 063469R1 (PLACNOB01), 215814F1 (STOMNOT01), 257941R6 (HNT2RAT01), 415908X20F1 (BRSTNOT01), 678003X11 (CRBLNOT01), 678003X17 (CRBLNOT01), 2731157F6 (OVARTUT04)

Table 2

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
1	192	S35 T45	N14	Transmembrane domain: F173-I191 zinc finger C3HC4 (RING) domain: Q39-K87	RMA1 RING zinc finger protein [Arabidopsis thaliana] g3164222	BLAST-Genbank MOTIFS HMMER-PFAM PROFILES-SCAN
2	169	S8 S10 T31 S51		zinc finger C2H2 domain: C97-H117	ZID zinc finger protein [Homo sapiens] g558599	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS
3	498	S49 S113 S123 T140 S170 S174 T212 S244 T313	N262	Leucine zipper: L278-L299 Wilm's tumor protein domain: T138-P152		MOTIFS BLIMPS-PRINTS
4	615	S13 S63 T126 S220 S414 S508 S86 S186 T406 S602	N124 N273 N546		DNA repair/ transcription protein Mms19p [S. cerevisiae] g1737175	BLAST-Genbank MOTIFS
5	120	S9 T18 S94	N40 N101	KRAB box domain: Q5-K72	zinc finger protein [Homo sapiens] g5738547	BLAST-Genbank MOTIFS BLAST-DOMO

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
6	543	S543 S21 S29 S78 T93 S95 S135 S158 S159 S160 T166 S168 S200 T202 S211 S213 T446 T503 S504 S509 S66 T231 S269 T329 S433 S436 S485 T488	N209	Zinc finger CCHC domain: I241-L258	Cellular nucleic acid binding protein [Gallus gallus] g2232217	BLAST-Genbank MOTIFS HMMER-PFAM
7	633	T77 S146 S153 S362 T449 T32 S126 T128 T132 S275 S315 S412 S538 S555 S559 S567 S573 S578 S580 S3 T4 T102 S115 S141 S174 S215 T286 S289 S358 S430 S436 T439 S442 T467 S475 S488 S492 S522 S620 S621 S628	N138 N171 N181 N213 N459 N487		Helicase II [Homo sapiens] g606833	BLAST-Genbank MOTIFS
8	312	S14 S37 T58 T139 T42 S94 S109 S115 S156 S198 T204 S285 S70 S99	N150 N294	Leucine zipper: L248-L269		MOTIFS
9	377	T339 T9 T11 T111 S142 T175 T363 S364 T48 S242 T300 T322 Y330	N288 N316 N361	Signal peptide: M1-S32 POZ domain: V18-W141	BTB domain (zinc finger) protein [C. elegans] g3876900	BLAST-Genbank MOTIFS SPSCAN BLAST-DOMO
10	170	S15 T42 S15 S82 S126		Zinc finger C2H2 domains: C36-H57 C73-H93 C114-H134 C145-H165	zinc finger protein [Xenopus laevis] g453468	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
11	160	T12 S132 T7 S144		Histone H3 signature: P17-E135	histone H3.1-I [Mus musculus] g1458132	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS
12	219	S36 S16 S26 S36 S98 T104 S181 Y45	N93		SNF7, transcrip- tional regulator [S. cerevisiae] g730759 (P39929)	BLAST- SwissProt MOTIFS
13	142	T4 S24		Zinc finger C2H2 domain: C70-H90		BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS
14	524	T466 T20 S183 T251 T274 T326 T334 S364 S482 S484 T145 T168 T197 S235 T458 S468 T471		Leucine zipper: L30-L51	Chromosome condensation protein XCAP-G [Xenopus laevis] g4191596	BLAST-Genbank MOTIFS

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
15	500	T10 S243 T244 T283 T473 S488 S61 T71 T92 T143 S178 S186 S297 T307 T439 S460 S5 T42 S75 T91 T102 T105 S160 S230 S238 S243 S277 S278 T283 S290 S327 S330 S335 S400 T429	N59 N398	ATP/GTP binding site (P-loop): G257-S264	Suppressor of yeast mitotic catastrophe [Xenopus laevis] g464003	BLAST-Genbank MOTIFS
16	119	S5 S103 T83	N2	Zinc finger C2H2 domain: C33-H49	Zfp64 (zinc finger protein) [Mus musculus] g1842216	BLAST-Genbank MOTIFS BLIMPS-BLOCKS
17	544	S13 S33 S186 S257 S329 S22 S40 S51 S81 S107 T108 S131 S154 S219 S264 S271 S286 S332 T338 T359 T9 T29 S69 T113	N86 N95 N105 N408 N411 N421		Tat-SF1 (cofactor for transcription elongation by HIV-1 Tat) [Homo sapiens] g2808420	BLAST-Genbank MOTIFS
18	869	T514 S55 S74 S118 T119 S198 S260 S445 T459 T170 T289 S438 S439 S697 S722 T783 Y185	N405 N418 N806		MHC class II transactivator CIITA form IV [Mus musculus] g3335110	BLAST-Genbank MOTIFS
19	128	T101 T25 S65 S83 T8 T25 T58		Zinc finger domain: F17-V55 KRAB box domain: Q12-Q79	KID2 (zinc finger) [Mus musculus] g6007771	BLAST-Genbank MOTIFS BLAST-PRODOM BLAST-DOMO

Table 2 (cont.)

Protein SEQ ID No:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
20	301	S179 T209 S25 T36 S53 T100 S115 T287 S288 S167 T225 T247 Y255	N213 N285		BTB domain (zinc finger) protein [C. elegans] g3876900	BLAST-Genbank MOTIFS
21	402	S101 T269 S311 T348 T24 T315 S364	N58 N117 N130 N152 N228 N309	Ankyrin repeat domain: G333-S364	BCL-3 proto-oncogene [Homo sapiens] g533381	BLAST-Genbank MOTIFS HMMER-PFAM
22	254	S118 T193 T201 T245 S80 S112 S206 S244	N2	Zinc finger C3HC4 (RING) domain: K22-G91	RING finger protein [Homo sapiens] g3462505	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS PROFILES CAN
23	553	S520 S6 S47 S70 S76 T77 T89 S91 S110 T114 S249 S340 S431 S441 S442 S446 S156 S234 T318 S401 S471 S523 S524 S530 T544 S548	N46 N108 N246	Signal Peptide: M1-A27 Linker histone signature: P157-P414	Chromatin- binding protein [Mus musculus] g1480112	BLAST-Genbank MOTIFS HMM-PFAM BLIMPS-PRINTS SPSCAN
24	461	S190 T15 T24 T58 T83 T89 S125 S98 T153 S166 S194 S237 S275 S303 T404 Y169	N279 N444	Zinc finger C2H2 domains: C267-H287 C295-H315 C323-H343 C351-H371 C379-H399 C407-H427 KRAB box domain: V14-P85	Zinc finger protein 10 [Homo sapiens] g3970712	BLAST-Genbank MOTIFS HMMER-PFAM BLAST-DOMO

Table 2 (cont.)

Protein SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
25	159	T18 S33 S38 S43 S44 S74 S101 S123	N121		Transcription elongation factor A SII homolog [Homo sapiens] g4336506	BLAST-Genbank MOTIFS
26	373	S273 T63 T124 S246 S304 S329 S145 S314	N80 N327	Zinc finger C2H2 domains: C150-H170 C178-H198 C206-H226 C234-H254 C262-H282 C317-H337 C345-H365	Zinc finger protein zfp47 [Homo sapiens] g1613858	BLAST-Genbank MOTIFS HMMER-PFAM
27	330	T115 S71 T118 T137 S225 S65 T148 T197 S284	N69 N145	Fork head domain signature: K18-R113 Signal peptide: M1-G36	Forkhead-related transcription factor FREAC-10 [Homo sapiens] g2829129	BLAST-Genbank MOTIFS HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS SPSCAN
28	396	S153 S6 S128 S138 T152 T314 S363 S377 S12 S85 S91 S94 T104 T238 Y349	N335	Fork head associated domain: Y281-G361	Transcription termination factor Rho [Micrococcus luteus] g1666540	BLAST-Genbank MOTIFS HMMER-PFAM
29	126	T5 S14 S48 T54		KRAB box domain: M1-V71	Zinc finger protein [Homo sapiens] g498723	BLAST-Genbank MOTIFS BLAST-DOMO

Table 2 (cont.)

Protein, SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosyla- tion Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
30	519	T92 T344 S357 S144 S174 T362 S431 T433 T443 S484 S507 T57 T237 T352 S383 T458 S461 S470 Y316 Y393	N272 N346	Zinc finger C2H2 domain: P69-S82	Transcription factor IIC63 [Homo sapiens] g5281316	BLAST-Genbank MOTIFS BLIMPS-PRINTS
31	493	S61 S91 S132 S161 S225 S306 T323 T404 T413 S484 S488 S121 T148 T153 S335 T369 S418		Signal peptide: M1-C41 dsRNA-binding motif: P370-Q434	Histone H4 [Entamoeba histolytica] g642230	BLAST-Genbank MOTIFS HMMER-PFAM SPSCAN
32	516	T81 T62 S299 S327 S467 Y499	N48 N160 N441	Zinc finger C2H2 domains: C123-H143 C151-H171 C179-H199 C207-H227 C235-H255 C263-H283 C291-H311 C319-H339 C347-H367 C375-H395 C403-H423 C431-H451 C459-H479 C487-H507	b34I8.1 (Kruppel related zinc finger protein 184) [Homo sapiens] g3135968	BLAST-Genbank MOTIFS HMMER-PFAM

Table 3

Polynucleotide SEQ ID NO:	Selected Fragments of Nucleic Acid Sequence	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
33	205-249	Reproductive (0.292) Nervous (0.146) Gastrointestinal (0.135)	Cell proliferation (0.594) Inflammation (0.333)	PBLUESCRIPT
34	702-746	Reproductive (0.294) Gastrointestinal (0.206) Nervous (0.176)	Cell proliferation (0.765) Inflammation (0.265)	PSPORT1
35	336-380	Hematopoietic/Immune (0.320) Reproductive (0.200) Gastrointestinal (0.160) Nervous (0.160)	Cell proliferation (0.680) Inflammation (0.360)	PSPORT1
36	1514-1558 1946-1990	Nervous (0.238) Reproductive (0.214) Hematopoietic/Immune (0.155)	Cell proliferation (0.560) Inflammation (0.405)	pINCY
37	619-663	Nervous (0.364) Reproductive (0.273) Hematopoietic/Immune (0.189)	Cell proliferation (0.636) Inflammation (0.364)	pINCY
38	865-909	Hematopoietic/Immune (0.200) Gastrointestinal (0.160) Nervous (0.160) Reproductive (0.160)	Cell proliferation (0.560) Inflammation (0.360)	pINCY
39	541-585	Reproductive (0.269) Hematopoietic/Immune (0.212) Nervous (0.212)	Cell proliferation (0.714) Inflammation (0.384)	pINCY
40	488-532	Reproductive (0.212) Nervous (0.182) Gastrointestinal (0.121) Hematopoietic/Immune (0.121)	Cell proliferation (0.636) Inflammation (0.303)	pINCY

Table 3 (cont.)

Polynucleotide SEQ ID NO:	Selected Fragments of Nucleic Acid Sequence	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
41	227-271	Nervous (0.248) Reproductive (0.239) Gastrointestinal (0.110)	Cell proliferation (0.651) Inflammation (0.303)	pINCY
42	261-305	Nervous (0.222) Reproductive (0.148) Cardiovascular (0.148) Gastrointestinal (0.148)	Cell proliferation (0.407) Inflammation (0.481)	pINCY
43	460-504	Reproductive (0.667) Hematopoietic/Immune (0.333)	Cell proliferation (0.333) Inflammation (0.667)	pINCY
44	664-708	Reproductive (0.270) Nervous (0.184) Cardiovascular (0.132)	Cell proliferation (0.691) Inflammation (0.289)	PSPORT1
45	139-183	Reproductive (0.310) Nervous (0.172) Gastrointestinal (0.172)	Cell proliferation (0.793) Inflammation (0.241)	PSPORT1
46	272-316	Hematopoietic/Immune (0.348) Nervous (0.174) Cardiovascular (0.087) Gastrointestinal (0.087) Musculoskeletal (0.087) Reproductive (0.087)	Cell proliferation (0.565) Inflammation (0.348)	PBLUESCRIPT
47	163-207 604-648	Cardiovascular (0.200) Reproductive (0.200) Nervous (0.140)	Cell proliferation (0.520) Inflammation (0.380)	PBLUESCRIPT
48	434-488	Cardiovascular (0.200) Endocrine (0.200) Hematopoietic/Immune (0.200) Nervous (0.200) Reproductive (0.200)	Cell proliferation (0.600) Inflammation (0.200)	PSPORT1

Table 3 (cont.)

Polynucleotide SEQ ID NO:	Selected Fragments of Nucleic Acid Sequence	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
49	389-433 1028-1072	Nervous (0.235) Reproductive (0.235) Hematopoietic/Immune (0.157)	Cell proliferation (0.491) Inflammation (0.353)	PSPORT1
50	650-694 2000-2044	Cardiovascular (0.292) Nervous (0.208) Gastrointestinal (0.167)	Cell proliferation (0.625) Inflammation (0.292)	pINCY
51	566-610	Reproductive (0.500) Gastrointestinal (0.250) Nervous (0.125) Urologic (0.125)	Cell proliferation (0.250) Inflammation (0.375)	pINCY
52	218-262	Reproductive (0.205) Cardiovascular (0.182) Gastrointestinal (0.182) Nervous (0.182)	Cell proliferation (0.409) Inflammation (0.318)	pINCY
53	51-95	Hematopoietic/Immune (0.242) Reproductive (0.212) Gastrointestinal (0.182)	Cell proliferation (0.545) Inflammation (0.454)	pINCY
54	1008-1052	Cardiovascular (0.556) Endocrine (0.111) Gastrointestinal (0.111) Musculoskeletal (0.111) Reproductive (0.111)	Cell proliferation (0.333) Inflammation (0.556)	pINCY
55	302-346	Reproductive (0.220) Hematopoietic/Immune (0.200) Nervous (0.200)	Cell proliferation (0.560) Inflammation (0.360)	pINCY
56	801-845	Hematopoietic/Immune (0.500) Reproductive (0.333) Cardiovascular (0.167)	Cell proliferation (0.667) Inflammation (0.500)	pINCY

Table 3 (cont.)

Polynucleotide SEQ ID NO:	Selected Fragments of Nucleic Acid Sequence	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
57	619-663	Reproductive (0.377) Nervous (0.188) Cardiovascular (0.116)	Cell proliferation (0.594) Inflammation (0.304)	pINCY
58	121-165	Hematopoietic/Immune (0.385) Cardiovascular (0.154) Dermatologic (0.154) Reproductive (0.154)	Cell proliferation (0.615) Inflammation (0.461)	pINCY
59	444-488 813-857	Cardiovascular (0.357) Reproductive (0.286) Nervous (0.143)	Cell proliferation (0.714) Inflammation (0.143)	pINCY
60	471-515 1146-1190	Reproductive (0.357) Hematopoietic/Immune (0.286) Cardiovascular (0.143) Gastrointestinal (0.143)	Cell proliferation (0.500) Inflammation (0.429)	pINCY
61	397-443	Nervous (0.320) Hematopoietic/Immune (0.200) Cardiovascular (0.160) Reproductive (0.160)	Cell proliferation (0.560) Inflammation (0.440)	pINCY
62	1703-1747	Reproductive (0.238) Nervous (0.168) Hematopoietic/Immune (0.139)	Cell proliferation (0.653) Inflammation (0.208)	pINCY
63	928-972 1297-1341	Reproductive (0.259) Gastrointestinal (0.222) Nervous (0.185)	Cell proliferation (0.519) Inflammation (0.407)	pINCY
64	647-691	Reproductive (0.314) Nervous (0.186) Gastrointestinal (0.127)	Cell proliferation (0.588) Inflammation (0.362)	pINCY

Table 4

Polynucleotide SEQ ID NO:	Library	Library Comment
33	HYPONOB01	This library was constructed using RNA (Clontech, #6579-2) isolated from the hypothalamus tissues of 51 male and female Caucasian donors, 16 to 75 years old.
34	LUNGNOT04	This library was constructed using RNA isolated from the lung tissue of a 2-year-old Hispanic male, who died from cerebral anoxia.
35	BRSTTUT03	This library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.
36	MENITUT03	This library was constructed using RNA isolated from brain meningioma tissue removed from a 35-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated a benign neoplasm in the right cerebellopontine angle of the brain. Patient history included hypothyroidism. Family history included myocardial infarction and breast cancer.
37	BRSTNOT07	This library was constructed using RNA isolated from diseased breast tissue removed from a 43-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated mildly proliferative fibrocystic changes with epithelial hyperplasia, papillomatosis, and duct ectasia. Pathology for the associated tumor tissue indicated invasive grade 4, nuclear grade 3 mammary adenocarcinoma with extensive comedo necrosis. Family history included epilepsy, cardiovascular disease, and type II diabetes.
38	THYRNOT03	This library was constructed using RNA isolated from thyroid tissue removed from the left thyroid of a 28-year-old Caucasian female during a complete thyroidectomy. Pathology indicated a small nodule of adenomatous hyperplasia present in the left thyroid. Pathology for the associated tumor tissue indicated dominant follicular adenoma, forming a well-encapsulated mass in the left thyroid.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
39	CORPNOT02	This library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
40	BRAITUT07	This library was constructed using RNA isolated from left frontal lobe tumor tissue removed from the brain of a 32-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated low grade desmoplastic neuronal neoplasm, type not otherwise specified. The lesion formed a firm, circumscribed cyst-associated mass involving white matter and cortex. Family history included atherosclerotic coronary artery disease.
41	BRAINOT14	This library was constructed using RNA isolated from brain tissue removed from the left frontal lobe of a 40-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated grade 4 gemistocytic astrocytoma.
42	COLNNOT22	This library was constructed using RNA isolated from colon tissue removed from a 56-year-old Caucasian female with Crohn's disease during a partial resection of the small intestine. Pathology indicated Crohn's disease of the ileum and ileal-colonic anastomosis, causing a fistula at the anastomotic site that extended into pericolonic fat. The ileal mucosa showed linear and punctate ulcers with intervening normal tissue. Family history included irritable bowel syndrome.
43	LEUKNOT02	This library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
44	BRSTNOT04	This library was constructed using RNA isolated from breast tissue removed from a 62-year-old East Indian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 ductal carcinoma. Patient history included benign hypertension, hyperlipidemia, and hematuria. Family history included cerebrovascular and cardiovascular disease, hyperlipidemia, and liver cancer.
45	LUNGAST01	This library was constructed using RNA isolated from the lung tissue of a 17-year-old Caucasian male, who died from head trauma. Patient history included asthma.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
46	TESTNOT03	This library was constructed using polyA RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
47	TESTNOT03	This library was constructed using polyA RNA isolated from testicular tissue removed from a 37-year-old Caucasian male, who died from liver disease. Patient history included cirrhosis, jaundice, and liver failure.
48	BRSTNOT05	This library was constructed using RNA isolated from breast tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated multicentric invasive grade 4 lobular carcinoma. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular and cardiovascular disease, breast and prostate cancer, and type I diabetes.
49	OVARNOT02	This library was constructed using RNA isolated from ovarian tissue removed from a 59-year-old Caucasian female who died of a myocardial infarction. Patient history included cardiomyopathy, coronary artery disease, previous myocardial infarctions, hypercholesterolemia, hypotension, and arthritis.
50	LUNGNOT20	This library was constructed using RNA isolated from right upper lobe lung tissue removed from a 61-year-old Caucasian male. Pathology indicated panacinal emphysema with blebs in the right anterior upper lobe and apex, as well as emphysema in the right posterior upper lobe. Patient history included angina pectoris, and gastric ulcer. Family history included a subdural hemorrhage, cancer of an unidentified site, atherosclerotic coronary artery disease, and pneumonia.
51	ISLTNOT01	This library was constructed using RNA isolated from a pooled collection of pancreatic islet cells.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
52	OVARTUT02	This library was constructed using RNA isolated from ovarian tumor tissue removed from a 51-year-old Caucasian female during an exploratory laparotomy, total abdominal hysterectomy, salpingo-oophorectomy, and an incidental appendectomy. Pathology indicated mucinous cystadenoma presenting as a multiloculated neoplasm involving the entire left ovary. The right ovary contained a follicular cyst and a hemorrhagic corpus luteum. The uterus showed proliferative endometrium and a single intramural leiomyoma. The peritoneal biopsy indicated benign glandular inclusions consistent with endosalpingiosis. Family history included atherosclerotic coronary artery disease, benign hypertension, breast cancer, and uterine cancer.
53	COLNTUT15	This library was constructed using RNA isolated from colon tumor tissue obtained from a 64-year-old Caucasian female during a right hemicolectomy with ileostomy and bilateral salpingo-oophorectomy (removal of the fallopian tubes and ovaries). Pathology indicated an invasive grade 3 adenocarcinoma. Patient history included hypothyroidism, depression, and anemia. Family history included colon cancer and uterine cancer.
54	BONTNOT01	This library was constructed using RNA isolated from tibial periosteum removed from a 20-year-old Caucasian male during a hemipelvectomy with amputation above the knee. Pathology for the associated tumor tissue indicated partially necrotic and cystic osteoblastic grade 3 osteosarcoma (post-chemotherapy). Family history included osteogenesis imperfecta, closed fracture, and type II diabetes.
55	LUNGTUT12	This library was constructed using RNA isolated from tumorous lung tissue removed from a 70-year-old Caucasian female during a lung lobectomy of the left upper lobe. Pathology indicated grade 3 (of 4) adenocarcinoma and vascular invasion. Patient history included tobacco abuse, depressive disorder, anxiety state, and skin cancer. Family history included cerebrovascular disease, congestive heart failure, colon cancer, depressive disorder, and primary liver cancer.
56	THYMFET02	This library was constructed using RNA isolated from thymus tissue removed from a Caucasian female fetus, who died at 17 weeks' gestation from anencephalus.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
57	CERVNOT03	This library was constructed using RNA isolated from uterine cervical tissue removed from a 40-year-old Caucasian female during a vaginal hysterectomy with bilateral salpingo-oophorectomy and dilation and curettage. Pathology indicated secretory phase endometrium.
58	TLYJINT01	This library was constructed using RNA isolated from a Jurkat cell line derived from the T cells of a male. The cells were treated for 18 hours with 50 ng/ml phorbol ester (PMA) and 1 micromolar calcium ionophore. Patient history included acute T-cell leukemia.
59	HEAANOT01	This library was constructed using RNA isolated from right coronary and right circumflex coronary artery tissue removed from the explanted heart of a 46-year-old Caucasian male during a heart transplantation. Patient history included myocardial infarction from total occlusion of the left anterior descending coronary artery, atherosclerotic coronary artery disease, hyperlipidemia, myocardial ischemia, dilated cardiomyopathy, left ventricular dysfunction, and tobacco abuse. Previous surgeries included cardiac catheterization. Family history included atherosclerotic coronary artery disease.
60	BRSTNOT25	This library was constructed using RNA isolated from breast tissue removed from a 35-year-old Caucasian female during a bilateral reduction mastoplasmy. Family history included uterine cancer, hyperlipidemia, benign hypertension, acute myocardial infarction, cerebrovascular disease, atherosclerotic coronary artery disease, and type II diabetes.
61	TLYMTXT01	This library was constructed using RNA isolated from activated allogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male. The cells were incubated for 6 hours in the presence of OKT3 mAb (1 microgram/ml OKT3), anti-CD28 mAb (1 ug/ml) and 5% human serum. The patient had no allergies.
62	BRAVTEXT03	This library was constructed using RNA isolated from treated astrocytes removed from the brain of a female fetus who died after 22 weeks' gestation. The cells were treated with tumor necrosis factor (TNF) alpha and interleukin 1 (IL-1), 10ng/ml each for 24 hours.

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Library	Library Comment
63	BRSTTMT02	This library was constructed using RNA isolated from diseased right breast tissue removed from a 46-year-old Caucasian female during a unilateral extended simple mastectomy and open breast biopsy. Pathology indicated mildly proliferative fibrocystic change, including intraductal duct ectasia, papilloma formation, and ductal hyperplasia. Pathology for the associated tumor tissue indicated multifocal ductal carcinoma in situ, both comedo and non-comedo types, nuclear grade 2 with extensive intraductal calcifications. Patient history included deficiency anemia, normal delivery, chronic sinusitis, extrinsic asthma, and kidney infection. Family history included type II diabetes, benign hypertension, cerebrovascular disease, skin cancer, and hyperlipidemia.
64	BRAUNOT01	This library was constructed using RNA isolated from caudate/putamen/nucleus accumbens tissue removed from the brain of a 35-year-old Caucasian male who died from cardiac failure. Pathology indicated moderate leptomenigeal fibrosis and multiple microinfarctions of the cerebral neocortex. Microscopically, the cerebral hemisphere revealed moderate fibrosis of the leptomeninges with focal calcifications. There was evidence of shrunken and slightly eosinophilic pyramidal neurons throughout the cerebral hemispheres. In addition, scattered throughout the cerebral cortex, there were multiple small microscopic areas of cavitation with surrounding gliosis. Patient history included dilated cardiomyopathy, congestive heart failure, cardiomegaly and an enlarged spleen and liver.

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
 - 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-32,
 - b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-32,
 - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32, and
 - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-32.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-32.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:33-64.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
 - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - 35 b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- 5 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64,
 b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33-64,
 c) a polynucleotide sequence complementary to a),
 d) a polynucleotide sequence complementary to b), and
10 e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization
20 complex is formed between said probe and said target polynucleotide or fragments thereof, and
 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

25

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
30 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A pharmaceutical composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

35

17. A pharmaceutical composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-32.

18. A method for treating a disease or condition associated with decreased expression of functional TXREG, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

10 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 b) detecting agonist activity in the sample.

20. A pharmaceutical composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

15

21. A method for treating a disease or condition associated with decreased expression of functional TXREG, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

20 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
 b) detecting antagonist activity in the sample.

23. A pharmaceutical composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

25

24. A method for treating a disease or condition associated with overexpression of functional TXREG, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 23.

30

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

 a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

35

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.

LAL, Preeti

YUE, Henry

TANG, Y. Tom

BAUGHN, Mariah R.

AZIMZAI, Yalda

TRAN, Bao

<120> HUMAN TRANSCRIPTIONAL REGULATOR PROTEINS

<130> PF-0713 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/140,109

<151> 1999-06-18

<160> 64

<170> PERL Program

<210> 1

<211> 192

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 091502CD1

<400> 1

Met	Ala	Ser	Lys	Gly	Pro	Ser	Ala	Ser	Ala	Ser	Pro	Glu	Asn	Ser
1				5					10					15
Ser	Ala	Gly	Gly	Pro	Ser	Gly	Ser	Ser	Asn	Gly	Ala	Gly	Glu	Ser
				20					25					30
Gly	Gly	Gln	Asp	Ser	Thr	Phe	Glu	Cys	Asn	Ile	Cys	Leu	Asp	Thr
				35					40					45
Ala	Lys	Asp	Ala	Val	Ile	Ser	Leu	Cys	Gly	His	Leu	Phe	Cys	Trp
				50					55					60
Pro	Cys	Leu	His	Gln	Trp	Leu	Glu	Thr	Arg	Pro	Asn	Arg	Gln	Val
				65					70					75
Cys	Pro	Val	Cys	Lys	Ala	Gly	Ile	Ser	Arg	Asp	Lys	Val	Ile	Pro
				80					85					90
Leu	Tyr	Gly	Arg	Gly	Ser	Thr	Gly	Gln	Gln	Asp	Pro	Arg	Glu	Lys
				95					100					105
Thr	Pro	Pro	Arg	Pro	Gln	Gly	Gln	Arg	Pro	Glu	Pro	Glu	Asn	Arg
				110					115					120
Gly	Gly	Phe	Gln	Gly	Phe	Gly	Phe	Gly	Asp	Gly	Gly	Phe	Gln	Met
				125					130					135
Ser	Phe	Gly	Ile	Gly	Ala	Phe	Pro	Phe	Gly	Ile	Phe	Ala	Thr	Ala
				140					145					150
Phe	Asn	Ile	Asn	Asp	Gly	Arg	Pro	Pro	Pro	Ala	Val	Pro	Gly	Thr
				155					160					165
Pro	Gln	Tyr	Val	Asp	Glu	Gln	Phe	Leu	Ser	Arg	Leu	Phe	Leu	Phe
				170					175					180
Val	Ala	Leu	Val	Ile	Met	Phe	Trp	Leu	Leu	Ile	Ala			
				185					190					

<210> 2

<211> 169

<212> PRT

<213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 763816CD1

<400> 2

Met	Asp	Ile	Lys	Gly	Gln	Glu	Ser	Ser	Ser	Asp	Gln	Glu	Gln	Val
1				5					10					15
Asp	Val	Glu	Ser	Ile	Asp	Phe	Ser	Lys	Glu	Asn	Lys	Met	Asp	Met
				20					25					30
Thr	Ser	Pro	Glu	Gln	Ser	Arg	Asn	Val	Leu	Gln	Phe	Thr	Glu	Glu
				35					40					45
Lys	Glu	Ala	Phe	Ile	Ser	Glu	Glu	Glu	Ile	Ala	Lys	Tyr	Met	Lys
				50					55					60
Arg	Gly	Lys	Gly	Lys	Tyr	Tyr	Cys	Lys	Ile	Cys	Cys	Cys	Arg	Ala
				65					70					75
Met	Lys	Lys	Gly	Ala	Val	Leu	His	His	Leu	Val	Asn	Lys	His	Asn
				80					85					90
Val	His	Ser	Pro	Tyr	Lys	Cys	Thr	Ile	Cys	Gly	Lys	Ala	Phe	Leu
				95					100					105
Leu	Glu	Ser	Leu	Leu	Lys	Asn	His	Val	Ala	Ala	His	Gly	Gln	Ser
				110					115					120
Leu	Leu	Lys	Cys	Pro	Arg	Cys	Asn	Phe	Glu	Ser	Asn	Phe	Pro	Arg
				125					130					135
Gly	Phe	Lys	Lys	His	Leu	Thr	His	Cys	Gln	Ser	Arg	His	Asn	Glu
				140					145					150
Glu	Ala	Asn	Lys	Lys	Leu	Met	Glu	Ala	Leu	Glu	Pro	Pro	Leu	Glu
				155					160					165
Glu	Gln	Gln	Ile											

<210> 3
 <211> 498
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 961184CD1

<400> 3

Met	Pro	Phe	Leu	Ser	Leu	His	Arg	Ser	Pro	His	Gly	Pro	Ser	Lys
1				5					10					15
Leu	Cys	Asp	Asp	Pro	Gln	Ala	Ser	Leu	Val	Pro	Glu	Pro	Val	Pro
				20					25					30
Gly	Gly	Cys	Gln	Glu	Pro	Glu	Glu	Met	Ser	Trp	Pro	Pro	Ser	Gly
				35					40					45
Glu	Ile	Ala	Ser	Pro	Pro	Glu	Leu	Pro	Ser	Ser	Pro	Pro	Pro	Gly
				50					55					60
Leu	Pro	Glu	Val	Ala	Pro	Asp	Ala	Thr	Ser	Thr	Gly	Leu	Pro	Asp
				65					70					75
Thr	Pro	Ala	Ala	Pro	Glu	Thr	Ser	Thr	Asn	Tyr	Pro	Val	Glu	Cys
				80					85					90
Thr	Glu	Gly	Ser	Ala	Gly	Pro	Gln	Ser	Leu	Pro	Leu	Pro	Ile	Leu
				95					100					105
Glu	Pro	Val	Lys	Asn	Pro	Cys	Ser	Val	Lys	Asp	Gln	Thr	Pro	Leu
				110					115					120
Gln	Leu	Ser	Val	Glu	Asp	Thr	Thr	Ser	Pro	Asn	Thr	Lys	Pro	Cys
				125					130					135
Pro	Pro	Thr	Pro	Thr	Thr	Pro	Glu	Thr	Ser	Pro	Pro	Pro	Pro	Pro
				140					145					150
Pro	Pro	Pro	Ser	Ser	Thr	Pro	Cys	Ser	Ala	His	Leu	Thr	Pro	Ser
				155					160					165
Ser	Leu	Phe	Pro	Ser	Ser	Leu	Glu	Ser	Ser	Ser	Glu	Gln	Lys	Phe
				170					175					180
Tyr	Asn	Phe	Val	Ile	Leu	His	Ala	Arg	Ala	Asp	Glu	His	Ile	Ala
				185					190					195
Leu	Arg	Val	Arg	Glu	Lys	Leu	Glu	Ala	Leu	Gly	Val	Pro	Asp	Gly
				200					205					210

Ala	Thr	Phe	Cys	Glu	Asp	Phe	Gln	Val	Pro	Gly	Arg	Gly	Glu	Leu
				215					220					225
Ser	Cys	Leu	Gln	Asp	Ala	Ile	Asp	His	Ser	Ala	Phe	Ile	Ile	Leu
				230					235					240
Leu	Leu	Thr	Ser	Asn	Phe	Asp	Cys	Arg	Leu	Ser	Leu	His	Gln	Val
				245					250					255
Asn	Gln	Ala	Met	Met	Ser	Asn	Leu	Thr	Arg	Gln	Gly	Ser	Pro	Asp
				260					265					270
Cys	Val	Ile	Pro	Phe	Leu	Pro	Leu	Glu	Ser	Ser	Pro	Ala	Gln	Leu
				275					280					285
Ser	Ser	Asp	Thr	Ala	Ser	Leu	Leu	Ser	Gly	Leu	Val	Arg	Leu	Asp
				290					295					300
Glu	His	Ser	Gln	Ile	Phe	Ala	Arg	Lys	Val	Ala	Asn	Thr	Phe	Lys
				305					310					315
Pro	His	Arg	Leu	Gln	Ala	Arg	Lys	Ala	Met	Trp	Arg	Lys	Glu	Gln
				320					325					330
Asp	Thr	Arg	Ala	Leu	Arg	Glu	Gln	Ser	Gln	His	Leu	Asp	Gly	Glu
				335					340					345
Arg	Met	Gln	Ala	Ala	Ala	Leu	Asn	Ala	Ala	Tyr	Ser	Ala	Tyr	Leu
				350					355					360
Gln	Ser	Tyr	Leu	Ser	Tyr	Gln	Ala	Gln	Met	Glu	Gln	Leu	Gln	Val
				365					370					375
Ala	Phe	Gly	Ser	His	Met	Ser	Phe	Gly	Thr	Gly	Ala	Pro	Tyr	Gly
				380					385					390
Ala	Arg	Met	Pro	Phe	Gly	Gly	Gln	Val	Pro	Leu	Gly	Ala	Pro	Pro
				395					400					405
Pro	Phe	Pro	Thr	Trp	Pro	Gly	Cys	Pro	Gln	Pro	Pro	Pro	Leu	His
				410					415					420
Ala	Trp	Gln	Ala	Gly	Thr	Pro	Pro	Pro	Pro	Ser	Pro	Gln	Pro	Ala
				425					430					435
Ala	Phe	Pro	Gln	Ser	Leu	Pro	Phe	Pro	Gln	Ser	Pro	Ala	Phe	Pro
				440					445					450
Thr	Ala	Ser	Pro	Ala	Pro	Pro	Gln	Ser	Pro	Gly	Leu	Gln	Pro	Leu
				455					460					465
Ile	Ile	His	His	Ala	Gln	Met	Val	Gln	Leu	Gly	Leu	Asn	Asn	His
				470					475					480
Met	Trp	Asn	Gln	Arg	Gly	Ser	Gln	Ala	Pro	Glu	Asp	Lys	Thr	Gln
				485					490					495
Glu	Ala	Glu												

<210> 4

<211> 615

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1255525CD1

<400> 4

Met	Leu	Leu	Gly	Phe	Leu	Lys	Leu	Gln	Gln	Lys	Trp	Ser	Tyr	Glu
1				5					10					15
Asp	Lys	Asp	Gln	Arg	Pro	Leu	Asn	Gly	Phe	Lys	Asp	Gln	Leu	Cys
				20					25					30
Ser	Leu	Val	Phe	Met	Ala	Leu	Thr	Asp	Pro	Ser	Thr	Gln	Leu	Gln
				35					40					45
Leu	Val	Gly	Ile	Arg	Thr	Leu	Thr	Val	Leu	Gly	Ala	Gln	Pro	Asp
				50					55					60
Leu	Leu	Ser	Tyr	Glu	Asp	Leu	Glu	Leu	Ala	Val	Gly	His	Leu	Tyr
				65					70					75
Arg	Leu	Ser	Phe	Leu	Lys	Glu	Asp	Ser	Gln	Ser	Cys	Arg	Val	Ala
				80					85					90
Ala	Leu	Glu	Ala	Ser	Gly	Thr	Leu	Ala	Ala	Leu	Tyr	Pro	Val	Ala
				95					100					105
Phe	Ser	Ser	His	Leu	Val	Pro	Lys	Leu	Ala	Glu	Glu	Leu	Arg	Val
				110					115					120
Gly	Glu	Ser	Asn	Leu	Thr	Asn	Gly	Asp	Glu	Pro	Thr	Gln	Cys	Ser

Arg His Leu Cys	125	Leu Gln Ala Leu	130	Ala Val Ser Thr	135
Pro Ser Ile Val	140	Glu Thr Leu Pro	145	Leu Leu Gln His	150
Trp Gln Val Asn	155	Arg Gly Asn Met	160	Ala Gln Ser Ser	165
Ile Ala Val Cys	170	Gln Ser Leu Arg	175	Met Ala Glu Lys	180
Gln Asp Pro Glu	185	Ser Cys Trp Tyr	190	Phe His Gln Thr	195
Cys Leu Leu Ala	200	Leu Ala Val Gln	205	Ala Ser Met Pro	210
Pro Ser Val Leu	215	Arg Lys Val Leu	220	Glu Asp Glu Val	225
Ala Met Val Ser	230	Val Ile Gly Thr	235	Ala Thr Thr His	240
Glu Leu Ala Ala	245	Gln Ser Val Thr	250	His Ile Val Pro	255
Asp Gly Asn Val	260	Ser Phe Leu Pro	265	Glu Asn Ser Phe	270
Phe Gln Pro Phe	275	Gln Asp Gly Ser	280	Ser Gly Gln Arg	285
Ala Leu Leu Met	290	Ala Phe Val Cys	295	Ser Leu Pro Arg	300
Ile Pro Gln Leu	305	Asn Gln Leu Met	310	Arg Glu Leu Leu	315
Cys Cys His Ser	320	Cys Pro Phe Ser	325	Ser Thr Ala Ala	330
Phe Ala Gly Leu	335	Leu Asn Lys His	340	Pro Ala Gly Gln	345
Glu Phe Leu Gln	350	Leu Ala Val Asp	355	Lys Val Glu Ala	360
Ser Gly Pro Cys	365	Arg Ser Gln Ala	370	Phe Thr Leu Leu	375
Thr Lys Ala Leu	380	Val Leu Arg Tyr	385	His Pro Leu Ser	390
Thr Ala Arg Leu	395	Met Gly Leu Leu	400	Ser Asp Pro Glu	405
Ala Ala Ala Asp	410	Gly Phe Ser Leu	415	Leu Met Ser Asp	420
Val Leu Thr Arg	425	Ala Gly His Ala	430	Glu Val Arg Ile	435
Gln Arg Phe Phe	440	Thr Asp Asn Val	445	Pro Ala Leu Val	450
His Ala Ala Pro	455	Gln Asp Val Lys	460	Pro Tyr Leu Lys	465
Ser His Val Leu	470	Asn Arg Leu Pro	475	Pro Val Leu Leu	480
Leu Pro Thr Leu	485	Leu Ser Leu Leu	490	Leu Glu Ala Leu	495
Asp Cys Val Val	500	Gln Leu Ser Thr	505	Leu Ser Cys Leu	510
Leu Leu Glu Ala	515	Pro Gln Val Met	520	Ser Leu His Val	525
Val Thr Lys Phe	530	Leu Asn Leu Ser	535	Thr Ser Pro Ser	540
Arg Ile Ala Ala	545	Leu Gln Cys Met	550	His Ala Leu Thr	555
Thr Pro Val Leu	560	Pro Tyr Lys Pro	565	Arg Ala Leu Thr	570
Ala Lys Pro Leu	575	Asp Asp Lys Lys	580	Arg Val Ile Arg	585
Val Ser Ala Arg	590	Gly Glu Trp Phe	595	Leu Val Arg Lys	600
	605		610	Gly Ser Pro Gly	615

<210> 5

<211> 120
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1297447CD1

<400> 5
 Met Ile Thr Ser Gln Gly Ser Val Ser Phe Arg Asp Val Thr Val
 1 5 10 15
 Gly Phe Thr Gln Glu Glu Trp Gln His Leu Asp Pro Ala Gln Arg
 20 25 30
 Thr Leu Tyr Arg Asp Val Met Leu Glu Asn Tyr Ser His Leu Val
 35 40 45
 Ser Val Gly Tyr Cys Ile Pro Lys Pro Glu Val Ile Leu Lys Leu
 50 55 60
 Glu Lys Gly Glu Glu Pro Trp Ile Leu Glu Glu Lys Phe Pro Ser
 65 70 75
 Gln Ser His Leu Gly Glu Leu Val Cys Ala Arg Trp Asn Leu Lys
 80 85 90
 Glu Gly Arg Ser Gln Arg Val Ser Leu Asp Asn Lys Thr Ile Glu
 95 100 105
 Met Phe Phe Arg Asn His Val Leu Glu Ala Pro Asp Leu Trp Lys
 110 115 120

<210> 6
 <211> 543
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1441094CD1

<400> 6
 Met Met Phe Gly Gly Tyr Glu Thr Ile Glu Ala Tyr Glu Asp Asp
 1 5 10 15
 Leu Tyr Arg Asp Glu Ser Ser Ser Glu Leu Ser Val Asp Ser Glu
 20 25 30
 Val Glu Phe Gln Leu Tyr Ser Gln Ile His Tyr Ala Gln Asp Leu
 35 40 45
 Asp Asp Val Ile Arg Glu Glu Glu His Glu Glu Lys Asn Ser Gly
 50 55 60
 Asn Ser Glu Ser Ser Ser Ser Lys Pro Asn Gln Lys Lys Leu Ile
 65 70 75
 Val Leu Ser Asp Ser Glu Val Ile Gln Leu Ser Asp Gly Ser Glu
 80 85 90
 Val Ile Thr Leu Ser Asp Glu Asp Ser Ile Tyr Arg Cys Lys Gly
 95 100 105
 Lys Asn Val Arg Val Gln Ala Gln Glu Asn Ala His Gly Leu Ser
 110 115 120
 Ser Ser Leu Gln Ser Asn Glu Leu Val Asp Lys Lys Cys Lys Ser
 125 130 135
 Asp Ile Glu Lys Pro Lys Ser Glu Glu Arg Ser Gly Val Ile Arg
 140 145 150
 Glu Val Met Ile Ile Glu Val Ser Ser Ser Glu Glu Glu Glu Ser
 155 160 165
 Thr Ile Ser Glu Gly Asp Asn Val Glu Ser Trp Met Leu Leu Gly
 170 175 180
 Cys Glu Val Asp Asp Lys Asp Asp Asp Ile Leu Leu Asn Leu Val
 185 190 195
 Gly Cys Glu Asn Ser Val Thr Glu Gly Glu Asp Gly Ile Asn Trp
 200 205 210
 Ser Ile Ser Asp Lys Asp Ile Glu Ala Gln Ile Ala Asn Asn Arg
 215 220 225
 Thr Pro Gly Arg Trp Thr Gln Arg Tyr Tyr Ser Ala Asn Lys Asn

Ile Ile Cys Arg	230	Asn Cys Asp Lys Arg	235	Gly His Leu Ser Lys Asn	240
Cys Pro Leu Pro	245	Arg Lys Val Arg Arg	250	Cys Phe Leu Cys Ser Arg	255
Arg Gly His Leu	260	Leu Tyr Ser Cys Pro	265	Ala Pro Leu Cys Glu Tyr	270
Cys Pro Val Pro	275	Lys Met Leu Asp His	280	Ser Cys Leu Phe Arg His	285
Ser Trp Asp Lys	290	Gln Cys Asp Arg Cys	295	His Met Leu Gly His Tyr	300
Thr Asp Ala Cys	305	Thr Glu Ile Trp Arg	310	Gln Tyr His Leu Thr Thr	315
Lys Pro Gly Pro	320	Pro Lys Lys Pro Lys	325	Thr Pro Ser Arg Pro Ser	330
Ala Leu Ala Tyr	335	Cys Tyr His Cys Ala	340	Gln Lys Gly His Tyr Gly	345
His Glu Cys Pro	350	Glu Arg Glu Val Tyr	355	Asp Pro Ser Pro Val Ser	360
Pro Phe Ile Cys	365	Tyr Tyr Asp Asp Lys	370	Tyr Glu Ile Gln Glu Arg	375
Glu Lys Arg Leu	380	Lys Gln Lys Ile Lys	385	Val Leu Lys Lys Asn Gly	390
Val Ile Pro Glu	395	Pro Ser Lys Leu Pro	400	Tyr Ile Lys Ala Ala Asn	405
Glu Asn Pro His	410	His Asp Ile Arg Lys	415	Gly Arg Ala Ser Trp Lys	420
Ser Asn Arg Trp	425	Pro Gln Glu Asn Lys	430	Glu Thr Gln Lys Glu Met	435
Lys Asn Lys Asn	440	Arg Asn Trp Glu Lys	445	His Arg Lys Ala Asp Arg	450
His Arg Glu Val	455	Asp Glu Asp Phe Pro	460	Arg Gly Pro Lys Thr Tyr	465
Ser Ser Pro Gly	470	Ser Phe Lys Thr Gln	475	Lys Pro Ser Lys Pro Phe	480
His Arg Ser Ser	485	His Tyr His Thr Ser	490	Arg Glu Asp Lys Ser Pro	495
Lys Glu Gly Lys	500	Arg Gly Lys Gln Lys	505	Lys Lys Glu Arg Cys Trp	510
Glu Asp Asp Asp	515	Asn Asp Asn Leu Phe	520	Leu Ile Lys Gln Arg Lys	525
Lys Lys Ser	530		535		540

<210> 7

<211> 633

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1479382CD1

<400> 7

Met Tyr Ser Thr Arg	1	Lys Asn Cys Ala Gln	10	Leu Trp Leu Gly Pro	15
Ala Ala Phe Ile Asn	5	His Asp Cys Arg Pro	20	Asn Cys Lys Phe Val	25
Ser Thr Gly Arg Asp	20	Thr Ala Cys Val Lys	25	Ala Leu Arg Asp Ile	30
Glu Pro Gly Glu Glu	35	Ile Ser Cys Tyr Tyr	40	Gly Asp Gly Phe Phe	45
Gly Glu Asn Asn Glu	50	Phe Cys Glu Cys Tyr	55	Thr Cys Glu Arg Arg	60
Gly Thr Gly Ala Phe	65	Lys Ser Arg Val Gly	70	Leu Pro Ala Pro Ala	75
Pro Val Ile Asn Ser	80	Lys Tyr Gly Leu Arg	85	Glu Thr Asp Lys Arg	90
	95		100		105

Leu	Asn	Arg	Leu	Lys	Lys	Leu	Gly	Asp	Ser	Ser	Lys	Asn	Ser	Asp
				110					115					120
Ser	Gln	Ser	Val	Ser	Ser	Asn	Thr	Asp	Ala	Asp	Thr	Thr	Gln	Glu
				125					130					135
Lys	Asn	Asn	Ala	Thr	Ser	Asn	Arg	Lys	Ser	Ser	Val	Gly	Val	Lys
				140					145					150
Lys	Asn	Ser	Lys	Ser	Arg	Thr	Leu	Thr	Arg	Gln	Ser	Met	Ser	Arg
				155					160					165
Ile	Pro	Ala	Ser	Ser	Asn	Ser	Thr	Ser	Ser	Lys	Leu	Thr	His	Ile
				170					175					180
Asn	Asn	Ser	Arg	Val	Pro	Lys	Lys	Leu	Lys	Lys	Pro	Ala	Lys	Pro
				185					190					195
Leu	Leu	Ser	Lys	Ile	Lys	Leu	Arg	Asn	His	Cys	Lys	Arg	Leu	Glu
				200					205					210
Gln	Lys	Asn	Ala	Ser	Arg	Lys	Leu	Glu	Met	Gly	Asn	Leu	Val	Leu
				215					220					225
Lys	Glu	Pro	Lys	Val	Val	Leu	Tyr	Lys	Asn	Leu	Pro	Ile	Lys	Lys
				230					235					240
Asp	Lys	Glu	Pro	Glu	Gly	Pro	Ala	Gln	Ala	Ala	Val	Ala	Ser	Gly
				245					250					255
Cys	Leu	Thr	Arg	His	Ala	Ala	Arg	Glu	His	Arg	Gln	Asn	Pro	Val
				260					265					270
Arg	Gly	Ala	His	Ser	Gln	Gly	Glu	Ser	Ser	Pro	Cys	Thr	Tyr	Ile
				275					280					285
Thr	Arg	Arg	Ser	Val	Arg	Thr	Arg	Thr	Asn	Leu	Lys	Glu	Ala	Ser
				290					295					300
Asp	Ile	Lys	Leu	Glu	Pro	Asn	Thr	Leu	Asn	Gly	Tyr	Lys	Ser	Ser
				305					310					315
Val	Thr	Glu	Pro	Cys	Pro	Asp	Ser	Gly	Glu	Gln	Leu	Gln	Pro	Ala
				320					325					330
Pro	Val	Leu	Gln	Glu	Glu	Glu	Leu	Ala	His	Glu	Thr	Ala	Gln	Lys
				335					340					345
Gly	Glu	Ala	Lys	Cys	His	Lys	Ser	Asp	Thr	Gly	Met	Ser	Lys	Lys
				350					355					360
Lys	Ser	Arg	Gln	Gly	Lys	Leu	Val	Lys	Gln	Phe	Ala	Lys	Ile	Glu
				365					370					375
Glu	Ser	Thr	Pro	Val	His	Asp	Ser	Pro	Gly	Lys	Asp	Asp	Ala	Val
				380					385					390
Pro	Asp	Leu	Met	Gly	Pro	His	Ser	Asp	Gln	Gly	Glu	His	Ser	Gly
				395					400					405
Thr	Val	Gly	Val	Pro	Val	Ser	Tyr	Thr	Asp	Cys	Ala	Pro	Ser	Pro
				410					415					420
Val	Gly	Cys	Ser	Val	Val	Thr	Ser	Asp	Ser	Phe	Lys	Thr	Lys	Asp
				425					430					435
Ser	Phe	Arg	Thr	Ala	Lys	Ser	Lys	Lys	Lys	Arg	Arg	Ile	Thr	Arg
				440					445					450
Tyr	Asp	Ala	Gln	Leu	Ile	Leu	Glu	Asn	Asn	Ser	Gly	Ile	Pro	Lys
				455					460					465
Leu	Thr	Leu	Arg	Arg	Arg	His	Asp	Ser	Ser	Ser	Lys	Thr	Asn	Asp
				470					475					480
Gln	Glu	Asn	Asp	Gly	Met	Asn	Ser	Ser	Lys	Ile	Ser	Ile	Lys	Leu
				485					490					495
Ser	Lys	Asp	His	Asp	Asn	Asp	Asn	Asn	Leu	Tyr	Val	Ala	Lys	Leu
				500					505					510
Asn	Asn	Gly	Phe	Asn	Ser	Gly	Ser	Gly	Ser	Ser	Ser	Thr	Lys	Leu
				515					520					525
Lys	Ile	Gln	Leu	Lys	Arg	Asp	Glu	Glu	Asn	Arg	Gly	Ser	Tyr	Thr
				530					535					540
Glu	Gly	Leu	His	Glu	Asn	Gly	Val	Cys	Cys	Ser	Asp	Pro	Leu	Ser
				545					550					555
Leu	Leu	Glu	Ser	Arg	Met	Glu	Val	Asp	Asp	Tyr	Ser	Gln	Tyr	Glu
				560					565					570
Glu	Glu	Ser	Thr	Asp	Asp	Ser	Ser	Ser	Ser	Glu	Gly	Asp	Glu	Glu
				575					580					585
Glu	Asp	Asp	Tyr	Asp	Asp	Asp	Phe	Glu	Asp	Asp	Phe	Ile	Pro	Leu
				590					595					600
Pro	Pro	Ala	Lys	Arg	Leu	Arg	Leu	Ile	Val	Gly	Lys	Asp	Ser	Ile

Asp	Ile	Asp	Ile	Ser	Ser	Arg	Arg	Arg	Glu	Asp	Gln	Ser	Leu	Arg
				605					610					615
Leu	Asn	Ala		620					625					630

<210> 8
 <211> 312
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1503131CD1

<400> 8

Met	Ser	Ala	Phe	Ser	Glu	Ala	Ala	Leu	Glu	Lys	Lys	Leu	Ser	Glu
1				5					10					15
Leu	Ser	Asn	Ser	Gln	Gln	Ser	Val	Gln	Thr	Leu	Ser	Leu	Trp	Leu
				20					25					30
Ile	His	His	Arg	Lys	His	Ser	Arg	Pro	Ile	Val	Thr	Val	Trp	Glu
				35					40					45
Arg	Glu	Leu	Arg	Lys	Ala	Lys	Pro	Asn	Arg	Lys	Leu	Thr	Phe	Leu
				50					55					60
Tyr	Leu	Ala	Asn	Asp	Val	Ile	Gln	Asn	Ser	Lys	Arg	Lys	Gly	Pro
				65					70					75
Glu	Phe	Thr	Lys	Asp	Phe	Ala	Pro	Val	Ile	Val	Glu	Ala	Phe	Lys
				80					85					90
His	Val	Ser	Ser	Glu	Thr	Asp	Glu	Ser	Cys	Lys	Lys	His	Leu	Gly
				95					100					105
Arg	Val	Leu	Ser	Ile	Trp	Glu	Glu	Arg	Ser	Val	Tyr	Glu	Asn	Asp
				110					115					120
Val	Leu	Glu	Gln	Leu	Lys	Gln	Ala	Leu	Tyr	Gly	Asp	Lys	Lys	Pro
				125					130					135
Arg	Lys	Arg	Thr	Tyr	Glu	Gln	Ile	Lys	Val	Asp	Glu	Asn	Glu	Asn
				140					145					150
Cys	Ser	Ser	Leu	Gly	Ser	Pro	Ser	Glu	Pro	Pro	Gln	Thr	Leu	Asp
				155					160					165
Leu	Val	Arg	Ala	Leu	Gln	Asp	Leu	Glu	Asn	Ala	Ala	Ser	Gly	Asp
				170					175					180
Ala	Ala	Val	His	Gln	Arg	Ile	Ala	Ser	Leu	Pro	Val	Glu	Val	Gln
				185					190					195
Glu	Val	Ser	Leu	Leu	Asp	Lys	Ile	Thr	Asp	Lys	Glu	Ser	Gly	Glu
				200					205					210
Arg	Leu	Ser	Lys	Met	Val	Glu	Asp	Ala	Cys	Met	Leu	Leu	Ala	Asp
				215					220					225
Tyr	Asn	Gly	Arg	Glu	Ala	Ala	Glu	Ile	Asp	Asp	Arg	Lys	Gln	Leu
				230					235					240
Thr	Arg	Met	Leu	Ala	Asp	Phe	Leu	Arg	Cys	Gln	Lys	Glu	Ala	Leu
				245					250					255
Ala	Glu	Lys	Glu	His	Lys	Leu	Glu	Glu	Tyr	Lys	Arg	Lys	Leu	Ala
				260					265					270
Arg	Val	Ser	Leu	Val	Arg	Lys	Glu	Leu	Arg	Ser	Arg	Ile	Gln	Ser
				275					280					285
Leu	Pro	Asp	Leu	Ser	Arg	Leu	Pro	Asn	Val	Thr	Gly	Ser	His	Met
				290					295					300
His	Leu	Pro	Phe	Ala	Gly	Asp	Ile	Tyr	Ser	Glu	Asp			
				305					310					

<210> 9
 <211> 377
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1594803CD1

<400> 9

```

Met Phe Asn Gly Gly Met Ala Thr Thr Ser Thr Glu Ile Glu Leu
 1      5      10      15
Pro Asp Val Glu Pro Ala Ala Phe Leu Ala Leu Leu Lys Phe Leu
 20      25      30
Tyr Ser Asp Glu Val Gln Ile Gly Pro Glu Thr Val Met Thr Thr
 35      40      45
Leu Tyr Thr Ala Lys Lys Tyr Ala Val Pro Ala Leu Glu Ala His
 50      55      60
Cys Val Glu Phe Leu Lys Lys Asn Leu Arg Ala Asp Asn Ala Phe
 65      70      75
Met Leu Leu Thr Gln Ala Arg Leu Phe Asp Glu Pro Gln Leu Ala
 80      85      90
Ser Leu Cys Leu Glu Asn Ile Asp Lys Asn Thr Ala Asp Ala Ile
 95     100     105
Thr Ala Glu Gly Phe Thr Asp Ile Asp Leu Asp Thr Leu Val Ala
110     115     120
Val Leu Glu Arg Asp Thr Leu Gly Ile Arg Glu Val Arg Leu Phe
125     130     135
Asn Ala Val Val Arg Trp Ser Glu Ala Glu Cys Gln Arg Gln Gln
140     145     150
Leu Gln Val Thr Pro Glu Asn Arg Arg Lys Val Leu Gly Lys Ala
155     160     165
Leu Gly Leu Ile Arg Phe Pro Leu Met Thr Ile Glu Glu Phe Ala
170     175     180
Ala Gly Pro Ala Gln Ser Gly Ile Leu Val Asp Arg Glu Val Val
185     190     195
Ser Leu Phe Leu His Phe Thr Val Asn Pro Lys Pro Arg Val Glu
200     205     210
Phe Ile Asp Arg Pro Arg Cys Cys Leu Arg Gly Lys Glu Cys Ser
215     220     225
Ile Asn Arg Phe Gln Gln Val Glu Ser Arg Trp Gly Tyr Ser Gly
230     235     240
Thr Ser Asp Arg Ile Arg Phe Ser Val Asn Lys Arg Ile Phe Val
245     250     255
Val Gly Phe Gly Leu Tyr Gly Ser Ile His Gly Pro Thr Asp Tyr
260     265     270
Gln Val Asn Ile Gln Ile Ile His Thr Asp Ser Asn Thr Val Leu
275     280     285
Gly Gln Asn Asp Thr Gly Phe Ser Cys Asp Gly Ser Ala Ser Thr
290     295     300
Phe Arg Val Met Phe Lys Glu Pro Val Glu Val Leu Pro Asn Val
305     310     315
Asn Tyr Thr Ala Cys Ala Thr Leu Lys Gly Pro Asp Ser His Tyr
320     325     330
Gly Thr Lys Gly Leu Arg Lys Val Thr His Glu Ser Pro Thr Thr
335     340     345
Gly Ala Lys Thr Cys Phe Thr Phe Cys Tyr Ala Ala Gly Asn Asn
350     355     360
Asn Gly Thr Ser Val Glu Asp Gly Gln Ile Pro Glu Val Ile Phe
365     370     375
Tyr Thr

```

<210> 10

<211> 170

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1736129CD1

<400> 10

```

Met Glu Asp Pro Asn Pro Glu Glu Asn Met Lys Gln Gln Asp Ser
 1      5      10      15
Pro Lys Glu Arg Ser Pro Gln Ser Pro Gly Gly Asn Ile Cys His
 20      25      30

```

```

Leu Gly Ala Pro Lys Cys Thr Arg Cys Leu Ile Thr Phe Ala Asp
      35      40      45
Ser Lys Phe Gln Glu Arg His Met Lys Arg Glu His Pro Ala Asp
      50      55      60
Phe Val Ala Gln Lys Leu Gln Gly Val Leu Phe Ile Cys Phe Thr
      65      70      75
Cys Ala Arg Ser Phe Pro Ser Ser Lys Ala Leu Ile Thr His Gln
      80      85      90
Arg Ser His Gly Pro Ala Ala Lys Pro Thr Leu Pro Val Ala Thr
      95     100     105
Thr Thr Ala Gln Pro Thr Phe Pro Cys Pro Asp Cys Gly Lys Thr
     110     115     120
Phe Gly Gln Ala Val Ser Leu Arg Arg His Arg Gln Met His Glu
     125     130     135
Val Arg Ala Pro Pro Gly Thr Phe Ala Cys Thr Glu Cys Gly Gln
     140     145     150
Asp Phe Ala Gln Glu Ala Gly Leu His Gln His Tyr Ile Arg His
     155     160     165
Ala Arg Gly Glu Leu
      170

```

<210> 11
 <211> 160
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1874312CD1

```

<400> 11
Met Ala Arg Thr Lys Gln Thr Ala Arg Lys Ser Thr Gly Gly Lys
  1      5      10      15
Ala Pro Arg Lys Gln Leu Ala Thr Lys Ala Ala Arg Lys Ser Ala
      20      25      30
Pro Ala Thr Gly Gly Val Lys Lys Pro His Arg Tyr Arg Pro Gly
      35      40      45
Thr Val Ala Leu Arg Glu Ile Arg Arg Tyr Gln Lys Ser Thr Glu
      50      55      60
Leu Leu Ile Arg Lys Leu Pro Phe Gln Arg Leu Val Arg Glu Ile
      65      70      75
Ala Gln Asp Phe Lys Thr Asp Leu Arg Phe Gln Ser Ser Ala Val
      80      85      90
Met Ala Leu Gln Glu Ala Cys Glu Ala Tyr Leu Val Gly Leu Phe
      95     100     105
Glu Asp Thr Asn Leu Cys Gly Ile Gln Arg Gln Ala Arg His Tyr
     110     115     120
His Ala Gln Gly His Pro Thr His Pro Pro Ala Ser Ala Glu Glu
     125     130     135
Arg Ala Val Ile Thr Val Gly Leu Ser Cys Arg Ser Lys Gln Arg
     140     145     150
Val Phe Phe Arg Ala Thr Thr Phe Ser Lys
     155     160

```

<210> 12
 <211> 219
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1969301CD1

```

<400> 12
Met Asn Arg Leu Phe Gly Lys Ala Lys Pro Lys Ala Pro Pro Pro
  1      5      10      15
Ser Leu Thr Asp Cys Ile Gly Thr Val Asp Ser Arg Ala Glu Ser

```

	20		25		30
Ile Asp Lys Lys	Ile Ser Arg Leu Asp	Ala Glu Leu Val Lys	Tyr		
	35		40		45
Lys Asp Gln Ile	Lys Lys Met Arg Glu Gly	Pro Ala Lys Asn Met			
	50		55		60
Val Lys Gln Lys	Ala Leu Arg Val Leu Lys	Gln Lys Arg Met Tyr			
	65		70		75
Glu Gln Gln Arg	Asn Leu Ala Gln Gln	Ser Phe Asn Met Glu			
	80		85		90
Gln Ala Asn Tyr	Thr Ile Gln Ser Leu Lys	Asp Thr Lys Thr Thr			
	95		100		105
Val Asp Ala Met	Lys Leu Gly Val Lys	Glu Met Lys Lys Ala Tyr			
	110		115		120
Lys Gln Val Lys	Ile Asp Gln Ile Glu Asp	Leu Gln Asp Gln Leu			
	125		130		135
Glu Asp Met Met	Glu Asp Ala Asn Glu Ile	Gln Glu Ala Leu Ser			
	140		145		150
Arg Ser Tyr Gly	Thr Pro Glu Leu Asp	Glu Asp Asp Leu Glu Ala			
	155		160		165
Glu Leu Asp Ala	Leu Gly Asp Glu Leu Leu	Ala Asp Glu Asp Ser			
	170		175		180
Ser Tyr Leu Asp	Glu Ala Ala Ser Ala Pro	Ala Ile Pro Glu Gly			
	185		190		195
Val Pro Thr Asp	Thr Lys Asn Lys Asp	Gly Val Leu Val Asp Glu			
	200		205		210
Phe Gly Leu Pro	Gln Ile Pro Ala Ser				
	215				

<210> 13
 <211> 142
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1986873CD1

<400> 13	
Met Asp Met Thr Ser	Pro Glu Gln Ser Arg Asn Val Leu Gln Phe
1	5
Thr Glu Glu Lys Glu	Ala Phe Ile Ser Glu Glu Glu Ile Ala Lys
	20
Tyr Met Lys Arg Gly	Lys Gly Lys Tyr Tyr Cys Lys Ile Cys Cys
	35
Cys Arg Ala Met Lys	Lys Gly Ala Val Leu His His Leu Val Asn
	50
Lys His Asn Val His	Ser Pro Tyr Lys Cys Thr Ile Cys Gly Lys
	65
Ala Phe Leu Leu Glu	Ser Leu Leu Lys Asn His Val Ala Ala His
	80
Gly Gln Ser Leu Leu	Lys Cys Pro Arg Cys Asn Phe Glu Ser Asn
	95
Phe Pro Arg Gly Phe	Lys Lys His Leu Thr His Cys Gln Ser Arg
	110
His Asn Glu Glu Ala	Asn Lys Lys Leu Met Glu Ala Leu Glu Pro
	125
Pro Leu Glu Glu Gln	Gln Ile
	140

<210> 14
 <211> 524
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2010820CD1

<400> 14

Met	Ala	Glu	Ile	Lys	Val	Lys	Leu	Ile	Glu	Ala	Lys	Glu	Ala	Leu
1				5					10					15
Glu	Asn	Cys	Ile	Thr	Leu	Gln	Asp	Phe	Asn	Arg	Ala	Ser	Glu	Leu
				20					25					30
Lys	Glu	Glu	Ile	Lys	Ala	Leu	Glu	Asp	Ala	Arg	Ile	Asn	Leu	Leu
				35					40					45
Lys	Glu	Thr	Glu	Gln	Leu	Glu	Ile	Lys	Glu	Val	His	Ile	Glu	Lys
				50					55					60
Asn	Asp	Ala	Glu	Thr	Leu	Gln	Lys	Cys	Leu	Ile	Leu	Cys	Tyr	Glu
				65					70					75
Leu	Leu	Lys	Gln	Met	Ser	Ile	Ser	Thr	Gly	Leu	Ser	Ala	Thr	Met
				80					85					90
Asn	Gly	Ile	Ile	Glu	Ser	Leu	Ile	Leu	Pro	Gly	Ile	Ile	Ser	Ile
				95					100					105
His	Pro	Val	Val	Arg	Asn	Leu	Ala	Val	Leu	Cys	Leu	Gly	Cys	Cys
				110					115					120
Gly	Leu	Gln	Asn	Gln	Asp	Phe	Ala	Arg	Lys	His	Phe	Val	Leu	Leu
				125					130					135
Leu	Gln	Val	Leu	Gln	Ile	Asp	Asp	Val	Thr	Ile	Lys	Ile	Ser	Ala
				140					145					150
Leu	Lys	Ala	Ile	Phe	Asp	Gln	Leu	Met	Thr	Phe	Gly	Ile	Glu	Pro
				155					160					165
Phe	Lys	Thr	Lys	Lys	Ile	Lys	Thr	Leu	His	Cys	Glu	Gly	Thr	Glu
				170					175					180
Ile	Asn	Ser	Asp	Asp	Glu	Gln	Glu	Ser	Lys	Glu	Val	Glu	Glu	Thr
				185					190					195
Ala	Thr	Ala	Lys	Asn	Val	Leu	Lys	Leu	Leu	Ser	Asp	Phe	Leu	Asp
				200					205					210
Ser	Glu	Val	Ser	Glu	Leu	Arg	Thr	Gly	Ala	Ala	Glu	Gly	Leu	Ala
				215					220					225
Lys	Leu	Met	Phe	Ser	Gly	Leu	Leu	Val	Ser	Ser	Arg	Ile	Leu	Ser
				230					235					240
Arg	Leu	Ile	Leu	Leu	Trp	Tyr	Asn	Pro	Val	Thr	Glu	Glu	Asp	Val
				245					250					255
Gln	Leu	Arg	His	Cys	Leu	Gly	Val	Phe	Phe	Pro	Val	Phe	Ala	Tyr
				260					265					270
Ala	Ser	Arg	Thr	Asn	Gln	Glu	Cys	Phe	Glu	Glu	Ala	Phe	Leu	Pro
				275					280					285
Thr	Leu	Gln	Thr	Leu	Ala	Asn	Ala	Pro	Ala	Ser	Ser	Pro	Leu	Ala
				290					295					300
Glu	Ile	Asp	Ile	Thr	Asn	Val	Ala	Glu	Leu	Leu	Val	Asp	Leu	Thr
				305					310					315
Arg	Pro	Ser	Gly	Leu	Asn	Pro	Gln	Ala	Lys	Thr	Ser	Gln	Asp	Tyr
				320					325					330
Gln	Ala	Leu	Thr	Val	His	Asp	Asn	Leu	Ala	Met	Lys	Ile	Cys	Asn
				335					340					345
Glu	Ile	Leu	Thr	Ser	Pro	Cys	Ser	Pro	Glu	Ile	Arg	Val	Tyr	Thr
				350					355					360
Lys	Ala	Leu	Ser	Ser	Leu	Glu	Leu	Ser	Ser	His	Leu	Ala	Lys	Asp
				365					370					375
Leu	Leu	Val	Leu	Leu	Asn	Glu	Ile	Leu	Glu	Gln	Val	Lys	Asp	Arg
				380					385					390
Thr	Cys	Leu	Arg	Ala	Leu	Glu	Lys	Ile	Lys	Ile	Gln	Leu	Glu	Lys
				395					400					405
Gly	Asn	Lys	Glu	Phe	Gly	Asp	Gln	Ala	Glu	Ala	Ala	Gln	Asp	Ala
				410					415					420
Thr	Leu	Thr	Thr	Thr	Thr	Phe	Gln	Asn	Glu	Asp	Glu	Lys	Asn	Lys
				425					430					435
Glu	Val	Tyr	Met	Thr	Pro	Leu	Arg	Gly	Val	Lys	Ala	Thr	Gln	Ala
				440					445					450
Ser	Lys	Ser	Thr	Gln	Leu	Lys	Thr	Asn	Arg	Gly	Gln	Arg	Lys	Val
				455					460					465
Thr	Val	Ser	Ala	Arg	Thr	Asn	Arg	Arg	Cys	Gln	Thr	Ala	Glu	Ala
				470					475					480
Asp	Ser	Glu	Ser	Asp	His	Glu	Val	Pro	Glu	Pro	Glu	Ser	Glu	Met
				485					490					495

Lys	Met	Arg	Leu	Pro	Arg	Arg	Ala	Lys	Thr	Ala	Ala	Leu	Glu	Lys
				500					505					510
Ser	Lys	Leu	Asn	Leu	Ala	Gln	Phe	Leu	Asn	Glu	Asp	Leu	Ser	
				515					520					

<210> 15
 <211> 500
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2013818CD1

<400> 15

Met	Pro	Gly	Gln	Ser	Val	Arg	Lys	Lys	Thr	Arg	Lys	Ala	Lys	Glu
1				5					10					15
Ile	Ser	Glu	Ala	Ser	Glu	Asn	Ile	Tyr	Ser	Asp	Val	Arg	Gly	Leu
				20					25					30
Ser	Gln	Asn	Gln	Gln	Ile	Pro	Gln	Asn	Ser	Val	Thr	Pro	Arg	Arg
				35					40					45
Gly	Arg	Arg	Lys	Lys	Glu	Val	Asn	Gln	Asp	Ile	Leu	Glu	Asn	Thr
				50					55					60
Ser	Ser	Val	Glu	Gln	Glu	Leu	Gln	Ile	Thr	Thr	Gly	Arg	Glu	Ser
				65					70					75
Lys	Arg	Leu	Lys	Ser	Ser	Gln	Leu	Leu	Glu	Pro	Ala	Val	Glu	Glu
				80					85					90
Thr	Thr	Lys	Lys	Glu	Val	Lys	Val	Ser	Ser	Val	Thr	Lys	Arg	Thr
				95					100					105
Pro	Arg	Arg	Ile	Lys	Arg	Ser	Val	Glu	Asn	Gln	Glu	Ser	Val	Glu
				110					115					120
Ile	Ile	Asn	Asp	Leu	Lys	Val	Ser	Thr	Val	Thr	Ser	Pro	Ser	Arg
				125					130					135
Met	Ile	Arg	Lys	Leu	Arg	Ser	Thr	Asn	Leu	Asp	Ala	Ser	Glu	Asn
				140					145					150
Thr	Gly	Asn	Lys	Gln	Asp	Asp	Lys	Ser	Ser	Asp	Lys	Gln	Leu	Arg
				155					160					165
Ile	Lys	His	Val	Arg	Arg	Val	Arg	Gly	Arg	Glu	Val	Ser	Pro	Ser
				170					175					180
Asp	Val	Arg	Glu	Asp	Ser	Asn	Leu	Glu	Ser	Ser	Gln	Leu	Thr	Val
				185					190					195
Gln	Ala	Glu	Phe	Asp	Met	Ser	Ala	Ile	Pro	Arg	Lys	Arg	Gly	Arg
				200					205					210
Pro	Arg	Lys	Ile	Asn	Pro	Ser	Glu	Asp	Val	Gly	Ser	Lys	Ala	Val
				215					220					225
Lys	Glu	Glu	Arg	Ser	Pro	Lys	Lys	Lys	Glu	Ala	Pro	Ser	Ile	Arg
				230					235					240
Arg	Arg	Ser	Thr	Arg	Asn	Thr	Pro	Ala	Lys	Ser	Glu	Asn	Val	Asp
				245					250					255
Val	Gly	Lys	Pro	Ala	Leu	Gly	Lys	Ser	Ile	Leu	Val	Pro	Asn	Glu
				260					265					270
Glu	Leu	Ser	Met	Val	Met	Ser	Ser	Lys	Lys	Lys	Leu	Thr	Lys	Lys
				275					280					285
Thr	Glu	Ser	Gln	Ser	Gln	Lys	Arg	Ser	Leu	His	Ser	Val	Ser	Glu
				290					295					300
Glu	Arg	Thr	Asp	Glu	Met	Thr	His	Lys	Glu	Thr	Asn	Glu	Gln	Glu
				305					310					315
Glu	Arg	Leu	Leu	Ala	Thr	Ala	Ser	Phe	Thr	Lys	Ser	Ser	Arg	Ser
				320					325					330
Ser	Arg	Thr	Arg	Ser	Ser	Lys	Ala	Ile	Leu	Leu	Pro	Asp	Leu	Ser
				335					340					345
Glu	Pro	Asn	Asn	Glu	Pro	Leu	Phe	Ser	Pro	Ala	Ser	Glu	Val	Pro
				350					355					360
Arg	Lys	Ala	Lys	Ala	Lys	Lys	Ile	Glu	Val	Pro	Ala	Gln	Leu	Lys
				365					370					375
Glu	Leu	Val	Ser	Asp	Leu	Ser	Ser	Gln	Phe	Val	Ile	Ser	Pro	Pro
				380					385					390

Ala	Leu	Arg	Ser	Arg	Gln	Lys	Asn	Thr	Ser	Asn	Lys	Asn	Lys	Leu	
				395					400					405	
Glu	Asp	Glu	Leu	Lys	Asp	Asp	Ala	Gln	Ser	Val	Glu	Thr	Leu	Gly	
				410					415					420	
Lys	Pro	Lys	Ala	Lys	Arg	Ile	Arg	Thr	Ser	Lys	Thr	Lys	Gln	Ala	
				425					430					435	
Ser	Lys	Asn	Thr	Glu	Lys	Glu	Ser	Ala	Trp	Ser	Pro	Pro	Pro	Ile	
				440					445					450	
Glu	Ile	Arg	Leu	Ile	Ser	Pro	Leu	Ala	Ser	Pro	Ala	Asp	Gly	Val	
				455					460					465	
Lys	Ser	Lys	Pro	Arg	Lys	Thr	Thr	Glu	Val	Thr	Gly	Thr	Gly	Leu	
				470					475					480	
Gly	Arg	Asn	Arg	Lys	Lys	Leu	Ser	Ser	Tyr	Pro	Lys	Gln	Ile	Leu	
				485					490					495	
Arg	Arg	Lys	Met	Leu											
				500											

<210> 16
 <211> 119
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2302032CD1

<400> 16															
Met	Asn	Ala	Ser	Ser	Glu	Gly	Glu	Ser	Phe	Ala	Gly	Ser	Val	Gln	
1				5					10					15	
Ile	Pro	Gly	Gly	Thr	Thr	Val	Leu	Val	Glu	Leu	Thr	Pro	Asp	Ile	
				20					25					30	
His	Ile	Cys	Gly	Ile	Cys	Lys	Gln	Gln	Phe	Asn	Asn	Leu	Asp	Ala	
				35					40					45	
Phe	Val	Ala	His	Lys	Gln	Ser	Gly	Cys	Gln	Leu	Thr	Gly	Thr	Ser	
				50					55					60	
Ala	Ala	Ala	Pro	Ser	Thr	Val	Gln	Phe	Val	Ser	Glu	Glu	Thr	Val	
				65					70					75	
Pro	Ala	Thr	Gln	Thr	Gln	Thr	Thr	Thr	Arg	Thr	Ile	Thr	Ser	Glu	
				80					85					90	
Thr	Gln	Thr	Ile	Thr	Gly	Thr	Ala	Gly	Ala	Trp	Gly	Ser	Arg	Pro	
				95					100					105	
Glu	Leu	Ala	Trp	Leu	Cys	Leu	Lys	His	Val	His	Gly	Thr	Cys		
				110					115						

<210> 17
 <211> 544
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2326109CD1

<400> 17															
Met	Ile	His	Val	Arg	Arg	His	Glu	Thr	Arg	Arg	Asn	Ser	Lys	Ser	
1				5					10					15	
His	Val	Pro	Glu	Gln	Lys	Ser	Arg	Val	Asp	Trp	Arg	Arg	Thr	Lys	
				20					25					30	
Arg	Ser	Ser	Ile	Ser	Gln	Leu	Leu	Asp	Ser	Asp	Glu	Glu	Leu	Asp	
				35					40					45	
Ser	Glu	Glu	Phe	Asp	Ser	Asp	Glu	Glu	Leu	Asp	Ser	Asp	Glu	Ser	
				50					55					60	
Phe	Glu	Asn	Asp	Glu	Glu	Leu	Asp	Ser	Asn	Lys	Gly	Pro	Asp	Cys	
				65					70					75	
Asn	Lys	Thr	Pro	Gly	Ser	Glu	Arg	Glu	Leu	Asn	Leu	Ser	Lys	Ile	
				80					85					90	
Gln	Ser	Glu	Gly	Asn	Asp	Ser	Lys	Cys	Leu	Ile	Asn	Ser	Gly	Asn	

	95		100		105
Gly Ser Thr Tyr	Glu Glu Glu Thr Asn	Lys Ile Lys His Arg	Asn		
	110		115		120
Ile Asp Leu Gln	Asp Gln Glu Lys His	Leu Ser Gln Glu Asp	Asn		
	125		130		135
Asp Leu Asn Lys	Gln Thr Gly Gln Ile	Ile Glu Asp Asp Gln	Glu		
	140		145		150
Lys His Leu Ser	Gln Glu Asp Asn Asp	Leu Asn Lys Gln Thr	Gly		
	155		160		165
Gln Ile Ile Glu	Asp Asp Leu Glu Glu	Glu Asp Ile Lys Arg	Gly		
	170		175		180
Lys Arg Lys Arg	Leu Ser Ser Val Met	Cys Asp Ser Asp Glu	Ser		
	185		190		195
Asp Asp Ser Asp	Ile Leu Val Arg Lys	Val Gly Val Lys Arg	Pro		
	200		205		210
Arg Arg Val Val	Glu Asp Glu Gly Ser	Ser Val Glu Met Glu	Gln		
	215		220		225
Lys Thr Pro Glu	Lys Thr Leu Ala Ala	Gln Lys Arg Glu Lys	Leu		
	230		235		240
Gln Lys Leu Lys	Glu Leu Ser Lys Gln	Arg Ser Arg Gln Arg	Arg		
	245		250		255
Ser Ser Gly Arg	Asp Phe Glu Asp Ser	Glu Lys Glu Ser Cys	Pro		
	260		265		270
Ser Ser Asp Glu	Val Asp Glu Glu Glu	Glu Glu Asp Asn Tyr	Glu		
	275		280		285
Ser Asp Glu Asp	Gly Asp Asp Tyr Ile	Ile Asp Asp Phe Val	Val		
	290		295		300
Gln Asp Glu Glu	Gly Asp Glu Glu Asn	Lys Asn Gln Gln Gly	Glu		
	305		310		315
Lys Leu Thr Thr	Ser Gln Leu Lys Leu	Val Lys Arg Asn Ser	Leu		
	320		325		330
Tyr Ser Phe Ser	Asp His Tyr Thr His	Phe Glu Arg Val Val	Lys		
	335		340		345
Ala Leu Leu Ile	Asn Ala Leu Asp Glu	Ser Phe Leu Gly Thr	Leu		
	350		355		360
Tyr Asp Gly Thr	Arg Gln Lys Ser Tyr	Ala Lys Asp Met Leu	Thr		
	365		370		375
Ser Leu His Tyr	Leu Asp Asn Arg Phe	Val Gln Pro Arg Leu	Glu		
	380		385		390
Ser Leu Val Ser	Arg Ser Arg Trp Lys	Glu Gln Tyr Lys Glu	Arg		
	395		400		405
Val Glu Asn Tyr	Ser Asn Val Ser Ile	His Leu Lys Asn Pro	Glu		
	410		415		420
Asn Cys Ser Cys	Gln Ala Cys Gly Leu	His Arg Tyr Cys Lys	Tyr		
	425		430		435
Ser Val His Leu	Ser Gly Glu Leu Tyr	Asn Thr Arg Thr Met	Gln		
	440		445		450
Ile Asp Asn Phe	Met Ser His Asp Lys	Gln Val Phe Thr Val	Gly		
	455		460		465
Arg Ile Cys Ala	Ser Arg Thr Arg Ile	Tyr His Lys Leu Lys	His		
	470		475		480
Phe Lys Phe Lys	Leu Tyr Gln Glu Cys	Cys Thr Ile Ala Met	Thr		
	485		490		495
Glu Glu Val Glu	Asp Glu Gln Val Lys	Glu Thr Val Glu Arg	Ile		
	500		505		510
Phe Arg Arg Ser	Lys Glu Asn Gly Trp	Ile Lys Glu Lys Tyr	Gly		
	515		520		525
Gln Leu Glu Glu	Tyr Leu Asn Phe Ala	Asp Tyr Phe Gln Glu	Glu		
	530		535		540
Lys Phe Glu Leu					

<210> 18

<211> 869

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2354751CD1

<400> 18

Met	Arg	Trp	Gly	His	His	Leu	Pro	Arg	Ala	Ser	Trp	Gly	Ser	Gly
1				5					10					15
Phe	Arg	Arg	Ala	Leu	Gln	Arg	Pro	Asp	Asp	Arg	Ile	Pro	Phe	Leu
				20					25					30
Ile	His	Trp	Ser	Trp	Pro	Leu	Gln	Gly	Glu	Arg	Pro	Phe	Gly	Pro
				35					40					45
Pro	Arg	Ala	Phe	Ile	Arg	His	His	Gly	Ser	Ser	Val	Asp	Ser	Ala
				50					55					60
Pro	Pro	Pro	Gly	Arg	His	Gly	Arg	Leu	Phe	Pro	Ser	Ala	Ser	Ala
				65					70					75
Thr	Glu	Ala	Ile	Gln	Arg	His	Arg	Arg	Asn	Leu	Ala	Glu	Trp	Phe
				80					85					90
Ser	Arg	Leu	Pro	Arg	Glu	Glu	Arg	Gln	Phe	Gly	Pro	Thr	Phe	Ala
				95					100					105
Leu	Asp	Thr	Val	His	Val	Asp	Pro	Val	Ile	Arg	Glu	Ser	Thr	Pro
				110					115					120
Asp	Glu	Leu	Leu	Arg	Pro	Pro	Ala	Glu	Leu	Ala	Leu	Glu	His	Gln
				125					130					135
Pro	Pro	Gln	Ala	Gly	Leu	Pro	Pro	Leu	Ala	Leu	Ser	Gln	Leu	Phe
				140					145					150
Asn	Pro	Asp	Ala	Cys	Gly	Arg	Arg	Val	Gln	Thr	Val	Val	Leu	Tyr
				155					160					165
Gly	Thr	Val	Gly	Thr	Gly	Lys	Ser	Thr	Leu	Val	Arg	Lys	Met	Val
				170					175					180
Leu	Asp	Trp	Cys	Tyr	Gly	Arg	Leu	Pro	Ala	Phe	Glu	Leu	Leu	Ile
				185					190					195
Pro	Phe	Ser	Cys	Glu	Asp	Leu	Ser	Ser	Leu	Gly	Pro	Ala	Pro	Ala
				200					205					210
Ser	Leu	Cys	Gln	Leu	Val	Ala	Gln	Arg	Tyr	Thr	Pro	Leu	Lys	Glu
				215					220					225
Val	Leu	Pro	Leu	Met	Ala	Ala	Ala	Gly	Ser	His	Leu	Leu	Phe	Val
				230					235					240
Leu	His	Gly	Leu	Glu	His	Leu	Asn	Leu	Asp	Phe	Arg	Leu	Ala	Gly
				245					250					255
Thr	Gly	Leu	Cys	Ser	Asp	Pro	Glu	Glu	Pro	Gln	Glu	Pro	Ala	Ala
				260					265					270
Ile	Ile	Val	Asn	Leu	Leu	Arg	Lys	Tyr	Met	Leu	Pro	Gln	Ala	Ser
				275					280					285
Ile	Leu	Val	Thr	Thr	Arg	Pro	Ser	Ala	Ile	Gly	Arg	Ile	Pro	Ser
				290					295					300
Lys	Tyr	Val	Gly	Arg	Tyr	Gly	Glu	Ile	Cys	Gly	Phe	Ser	Asp	Thr
				305					310					315
Asn	Leu	Gln	Lys	Leu	Tyr	Phe	Gln	Leu	Arg	Leu	Asn	Gln	Pro	Tyr
				320					325					330
Cys	Gly	Tyr	Ala	Val	Gly	Gly	Ser	Gly	Val	Ser	Ala	Thr	Pro	Ala
				335					340					345
Gln	Arg	Asp	His	Leu	Val	Gln	Met	Leu	Ser	Arg	Asn	Leu	Glu	Gly
				350					355					360
His	His	Gln	Ile	Ala	Ala	Ala	Cys	Phe	Leu	Pro	Ser	Tyr	Cys	Trp
				365					370					375
Leu	Val	Cys	Ala	Thr	Leu	His	Phe	Leu	His	Ala	Pro	Thr	Pro	Ala
				380					385					390
Gly	Gln	Thr	Leu	Thr	Ser	Ile	Tyr	Thr	Ser	Phe	Leu	Arg	Leu	Asn
				395					400					405
Phe	Ser	Gly	Glu	Thr	Leu	Asp	Ser	Thr	Asp	Pro	Ser	Asn	Leu	Ser
				410					415					420
Leu	Met	Ala	Tyr	Ala	Ala	Arg	Thr	Met	Gly	Lys	Leu	Ala	Tyr	Glu
				425					430					435
Gly	Val	Ser	Ser	Arg	Lys	Thr	Tyr	Phe	Ser	Glu	Glu	Asp	Val	Cys
				440					445					450
Gly	Cys	Leu	Glu	Ala	Gly	Ile	Arg	Thr	Glu	Glu	Glu	Phe	Gln	Leu
				455					460					465
Leu	His	Ile	Phe	Arg	Arg	Asp	Ala	Leu	Arg	Phe	Phe	Leu	Ala	Pro

	470		475		480
Cys Val Glu Pro	Gly Arg Ala Gly Thr	Phe Val Phe Thr Val	Pro		
	485		490		495
Ala Met Gln Glu	Tyr Leu Ala Ala Leu	Tyr Ile Val Leu Gly	Leu		
	500		505		510
Arg Lys Thr Thr	Leu Gln Lys Val Gly	Lys Glu Val Ala Glu	Leu		
	515		520		525
Val Gly Arg Val	Gly Glu Asp Val Ser	Leu Val Leu Gly Ile	Met		
	530		535		540
Ala Lys Leu Leu	Pro Leu Arg Ala Leu	Pro Leu Leu Phe Asn	Leu		
	545		550		555
Ile Lys Val Val	Pro Arg Val Phe Gly	Arg Met Val Gly Lys	Ser		
	560		565		570
Arg Glu Ala Val	Ala Gln Ala Met Val	Leu Glu Met Phe Arg	Glu		
	575		580		585
Glu Asp Tyr Tyr	Asn Asp Asp Val Leu	Asp Gln Met Gly Ala	Ser		
	590		595		600
Ile Leu Gly Val	Gly Gly Pro Arg Arg	His Pro Asp Glu Pro	Pro		
	605		610		615
Glu Asp Glu Val	Phe Glu Leu Phe Pro	Met Phe Met Gly Gly	Leu		
	620		625		630
Leu Ser Ala His	Asn Arg Ala Val Leu	Ala Gln Leu Gly Cys	Pro		
	635		640		645
Ile Lys Asn Leu	Asp Ala Leu Glu Asn	Ala Gln Ala Ile Lys	Lys		
	650		655		660
Lys Leu Gly Lys	Leu Gly Arg Gln Val	Leu Pro Pro Ser Glu	Leu		
	665		670		675
Leu Asp His Leu	Phe Phe His Tyr Glu	Phe Gln Asn Gln Arg	Phe		
	680		685		690
Ser Ala Glu Val	Leu Ser Ser Leu Arg	Gln Leu Asn Leu Ala	Gly		
	695		700		705
Val Arg Met Thr	Pro Val Lys Cys Thr	Val Val Ala Ala Val	Leu		
	710		715		720
Gly Ser Gly Arg	His Ala Leu Asp Glu	Val Asn Leu Ala Ser	Cys		
	725		730		735
Gln Leu Asp Pro	Ala Gly Leu Arg Thr	Leu Leu Pro Val Phe	Leu		
	740		745		750
Arg Ala Arg Lys	Leu Gly Leu Gln Leu	Asn Ser Leu Gly Pro	Glu		
	755		760		765
Ala Cys Lys Asp	Leu Arg Asp Leu Leu	Leu His Asp Gln Cys	Gln		
	770		775		780
Ile Thr Thr Leu	Arg Leu Ser Asn Asn	Pro Leu Thr Glu Ala	Gly		
	785		790		795
Val Ala Val Leu	Met Glu Gly Leu Ala	Gly Asn Thr Ser Val	Thr		
	800		805		810
His Leu Ser Leu	Leu His Thr Gly Leu	Gly Asp Glu Gly Leu	Glu		
	815		820		825
Leu Leu Ala Ala	Gln Leu Asp Arg Asn	Arg Gln Leu Gln Glu	Leu		
	830		835		840
Asn Val Ala Tyr	Asn Gly Ala Gly Asp	Thr Ala Ala Leu Ala	Leu		
	845		850		855
Ala Arg Ala Ala	Arg Glu His Pro Ser	Leu Glu Leu Leu Gln			
	860		865		

<210> 19

<211> 128

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2378058CD1

<400> 19

Met Met Thr Ala Val	Ser Leu Thr Thr	Arg Pro Gln Glu Ser	Val
1	5	10	15
Ala Phe Glu Asp Val	Ala Val Tyr Phe	Thr Thr Lys Glu Trp	Ala

Ile	Met	Val	Pro	Ala	Glu	Arg	Ala	Leu	Tyr	Arg	Asp	Val	Met	Leu
				20					25					30
Glu	Asn	Tyr	Glu	Ala	Val	Ala	Phe	Val	Val	Pro	Pro	Thr	Ser	Lys
				35					40					45
Pro	Ala	Leu	Val	Ser	His	Leu	Glu	Gln	Gly	Lys	Glu	Ser	Cys	Phe
				50					55					60
Thr	Gln	Pro	Gln	Gly	Val	Leu	Ser	Arg	Asn	Asp	Trp	Arg	Ala	Gly
				65					70					75
Trp	Ile	Gly	Tyr	Leu	Glu	Leu	Arg	Arg	Tyr	Thr	Tyr	Leu	Ala	Lys
				80					85					90
Ala	Val	Leu	Arg	Arg	Ile	Val	Ser	Lys	Ile	Phe	Arg	Asn	Arg	Gln
				95					100					105
Cys	Trp	Glu	Asp	Arg	Arg	Lys	Ala							120
				110					115					
				125										

<210> 20
 <211> 301
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2595747CD1

<400> 20

Met	Leu	Leu	Thr	Gln	Ala	Arg	Leu	Phe	Asp	Glu	Pro	Gln	Leu	Ala
1				5					10					15
Ser	Leu	Cys	Leu	Asp	Thr	Ile	Asp	Lys	Ser	Thr	Met	Asp	Ala	Ile
				20					25					30
Ser	Ala	Glu	Gly	Phe	Thr	Asp	Ile	Asp	Ile	Asp	Thr	Leu	Cys	Ala
				35					40					45
Val	Leu	Glu	Arg	Asp	Thr	Leu	Ser	Ile	Arg	Glu	Ser	Arg	Leu	Phe
				50					55					60
Gly	Ala	Val	Val	Arg	Trp	Ala	Glu	Ala	Glu	Cys	Gln	Arg	Gln	Gln
				65					70					75
Leu	Pro	Val	Thr	Phe	Gly	Asn	Lys	Gln	Lys	Val	Leu	Gly	Lys	Ala
				80					85					90
Leu	Ser	Leu	Ile	Arg	Phe	Pro	Leu	Met	Thr	Ile	Glu	Glu	Phe	Ala
				95					100					105
Ala	Gly	Pro	Ala	Gln	Ser	Gly	Ile	Leu	Ser	Asp	Arg	Glu	Val	Val
				110					115					120
Asn	Leu	Phe	Leu	His	Phe	Thr	Val	Asn	Pro	Lys	Pro	Arg	Val	Glu
				125					130					135
Tyr	Ile	Asp	Arg	Pro	Arg	Cys	Cys	Leu	Arg	Gly	Lys	Glu	Cys	Cys
				140					145					150
Ile	Asn	Arg	Phe	Gln	Gln	Val	Glu	Ser	Arg	Trp	Gly	Tyr	Ser	Gly
				155					160					165
Thr	Ser	Asp	Arg	Ile	Arg	Phe	Thr	Val	Asn	Arg	Arg	Ile	Ser	Ile
				170					175					180
Val	Gly	Phe	Gly	Leu	Tyr	Gly	Ser	Ile	His	Gly	Pro	Thr	Asp	Tyr
				185					190					195
Gln	Val	Asn	Ile	Gln	Ile	Ile	Glu	Tyr	Glu	Lys	Lys	Gln	Thr	Leu
				200					205					210
Gly	Gln	Asn	Asp	Thr	Gly	Phe	Ser	Cys	Asp	Gly	Thr	Ala	Asn	Thr
				215					220					225
Phe	Arg	Val	Met	Phe	Lys	Glu	Pro	Ile	Glu	Ile	Leu	Pro	Asn	Val
				230					235					240
Cys	Tyr	Thr	Ala	Cys	Ala	Thr	Leu	Lys	Gly	Pro	Asp	Ser	His	Tyr
				245					250					255
Gly	Thr	Lys	Gly	Leu	Lys	Lys	Val	Val	His	Glu	Thr	Pro	Ala	Ala
				260					265					270
Ser	Lys	Thr	Val	Phe	Phe	Phe	Phe	Ser	Ser	Pro	Gly	Asn	Asn	Asn
				275					280					285
Gly	Thr	Ser	Ile	Glu	Asp	Gly	Gln	Ile	Pro	Glu	Ile	Ile	Phe	Tyr
				290					295					300

Thr

```
<220>  
<221> misc_feature  
<223> Incyte ID No: 2634391CD1
```

Met	Ile	Asp	Gln	Ala	Ser	Leu	Tyr	Gln	Tyr	Ser	Pro	Gln	Asn	Gln
1				5					10					15
His	Val	Glu	Gln	Gln	Pro	His	Tyr	Thr	His	Lys	Pro	Thr	Leu	Glu
				20					25					30
Tyr	Ser	Pro	Phe	Pro	Ile	Pro	Pro	Gln	Ser	Pro	Ala	Tyr	Glu	Pro
				35					40					45
Asn	Leu	Phe	Asp	Gly	Pro	Glu	Ser	Gln	Phe	Cys	Pro	Asn	Gln	Ser
				50					55					60
Leu	Val	Ser	Leu	Leu	Gly	Asp	Gln	Arg	Glu	Ser	Glu	Asn	Ile	Ala
				65					70					75
Asn	Pro	Met	Gln	Thr	Ser	Ser	Ser	Val	Gln	Gln	Gln	Asn	Asp	Ala
				80					85					90
His	Leu	His	Ser	Phe	Ser	Met	Met	Pro	Ser	Ser	Ala	Cys	Glu	Ala
				95					100					105
Met	Val	Gly	His	Glu	Met	Ala	Ser	Asp	Ser	Ser	Asn	Thr	Ser	Leu
				110					115					120
Pro	Phe	Ser	Asn	Met	Gly	Asn	Pro	Met	Asn	Thr	Thr	Gln	Leu	Gly
				125					130					135
Lys	Ser	Leu	Phe	Gln	Trp	Gln	Val	Glu	Gln	Glu	Glu	Ser	Lys	Leu
				140					145					150
Ala	Asn	Ile	Ser	Gln	Asp	Gln	Phe	Leu	Ser	Lys	Asp	Ala	Asp	Gly
				155					160					165
Asp	Thr	Phe	Leu	His	Ile	Ala	Val	Ala	Gln	Gly	Arg	Arg	Ala	Leu
				170					175					180
Ser	Tyr	Val	Leu	Ala	Arg	Lys	Met	Asn	Ala	Leu	His	Met	Leu	Asp
				185					190					195
Ile	Lys	Glu	His	Asn	Gly	Gln	Ser	Ala	Phe	Gln	Val	Ala	Val	Ala
				200					205					210
Ala	Asn	Gln	His	Leu	Ile	Val	Gln	Asp	Leu	Val	Asn	Ile	Gly	Ala
				215					220					225
Gln	Val	Asn	Thr	Thr	Asp	Cys	Trp	Gly	Arg	Thr	Pro	Leu	His	Val
				230					235					240
Cys	Ala	Glu	Lys	Gly	His	Ser	Gln	Val	Leu	Gln	Ala	Ile	Gln	Lys
				245					250					255
Gly	Ala	Val	Gly	Ser	Asn	Gln	Phe	Val	Asp	Leu	Glu	Ala	Thr	Asn
				260					265					270
Tyr	Asp	Gly	Leu	Thr	Pro	Leu	His	Cys	Ala	Val	Ile	Ala	His	Asn
				275					280					285
Ala	Val	Val	His	Glu	Leu	Gln	Arg	Asn	Gln	Gln	Pro	His	Ser	Pro
				290					295					300
Glu	Val	Gln	Glu	Leu	Leu	Leu	Lys	Asn	Lys	Ser	Leu	Val	Asp	Thr
				305					310					315
Ile	Lys	Cys	Leu	Ile	Gln	Met	Gly	Ala	Ala	Val	Glu	Ala	Lys	Ala
				320					325					330
Tyr	Asn	Gly	Asn	Thr	Ala	Leu	His	Val	Ala	Ala	Ser	Leu	Gln	Tyr
				335					340					345
Arg	Leu	Thr	Gln	Leu	Asp	Ala	Val	Arg	Leu	Leu	Met	Arg	Lys	Gly

$$\begin{array}{ll} \langle 210 \rangle & 22 \\ \langle 211 \rangle & 254 \end{array}$$

<212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2637522CD1

<400> 22
 Met Asn Phe Thr Val Gly Phe Lys Pro Leu Leu Gly Asp Ala His
 1 5 10 15
 Ser Met Asp Asn Leu Glu Lys Gln Leu Ile Cys Pro Ile Cys Leu
 20 25 30
 Glu Met Phe Ser Lys Pro Val Val Ile Leu Pro Cys Gln His Asn
 35 40 45
 Leu Cys Arg Lys Cys Ala Asn Asp Val Phe Gln Ala Ser Asn Pro
 50 55 60
 Leu Trp Gln Ser Arg Gly Ser Thr Thr Val Ser Ser Gly Gly Arg
 65 70 75
 Phe Arg Cys Pro Ser Cys Arg His Glu Val Val Leu Asp Arg His
 80 85 90
 Gly Val Tyr Gly Leu Gln Arg Asn Leu Leu Val Glu Asn Ile Ile
 95 100 105
 Asp Ile Tyr Lys Gln Glu Ser Ser Arg Pro Leu His Ser Lys Ala
 110 115 120
 Glu Gln His Leu Met Cys Glu Glu His Glu Glu Glu Lys Ile Asn
 125 130 135
 Ile Tyr Cys Leu Ser Cys Glu Val Pro Thr Cys Ser Leu Cys Lys
 140 145 150
 Val Phe Gly Ala His Lys Asp Cys Glu Val Ala Pro Leu Pro Thr
 155 160 165
 Ile Tyr Lys Arg Gln Lys Ser Glu Leu Ser Asp Gly Ile Ala Met
 170 175 180
 Leu Val Ala Gly Asn Asp Arg Val Gln Ala Val Ile Thr Gln Met
 185 190 195
 Glu Glu Val Cys Gln Thr Ile Glu Asp Asn Ser Arg Arg Gln Lys
 200 205 210
 Gln Leu Leu Asn Gln Arg Phe Glu Ser Leu Cys Ala Val Leu Glu
 215 220 225
 Glu Arg Asn Gly Glu Leu Leu Gln Ala Leu Ala Arg Glu Gln Ala
 230 235 240
 Gly Gln Ala Ser Thr Arg Ser Asp Gly Thr His Ser Gly Gln
 245 250

<210> 23
 <211> 553
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2650980CD1

<400> 23
 Met Ala Thr Asp Thr Ser Gln Gly Glu Leu Val His Pro Lys Ala
 1 5 10 15
 Leu Pro Leu Ile Val Gly Ala Gln Leu Ile His Ala Asp Lys Leu
 20 25 30
 Gly Glu Lys Val Glu Asp Ser Thr Met Pro Ile Arg Arg Thr Val
 35 40 45
 Asn Ser Thr Arg Glu Thr Pro Pro Lys Ser Lys Leu Ala Glu Gly
 50 55 60
 Glu Glu Glu Lys Pro Glu Pro Asp Ile Ser Ser Glu Glu Ser Val
 65 70 75
 Ser Thr Val Glu Glu Gln Glu Asn Glu Thr Pro Pro Ala Thr Ser
 80 85 90
 Ser Glu Ala Glu Gln Pro Lys Gly Glu Pro Glu Asn Glu Glu Lys
 95 100 105

Glu	Glu	Asn	Lys	Ser	Ser	Glu	Glu	Thr	Lys	Lys	Asp	Glu	Lys	Asp
				110					115					120
Gln	Ser	Lys	Glu	Lys	Glu	Lys	Lys	Val	Lys	Lys	Thr	Ile	Pro	Ser
				125					130					135
Trp	Ala	Thr	Leu	Ser	Ala	Ser	Gln	Leu	Ala	Arg	Ala	Gln	Lys	Gln
				140					145					150
Thr	Pro	Met	Ala	Ser	Ser	Pro	Arg	Pro	Lys	Met	Asp	Ala	Ile	Leu
				155					160					165
Thr	Glu	Ala	Ile	Lys	Ala	Cys	Phe	Gln	Lys	Ser	Gly	Ala	Ser	Val
				170					175					180
Val	Ala	Ile	Arg	Lys	Tyr	Ile	Ile	His	Lys	Tyr	Pro	Ser	Leu	Glu
				185					190					195
Leu	Glu	Arg	Arg	Gly	Tyr	Leu	Leu	Lys	Gln	Ala	Leu	Lys	Arg	Glu
				200					205					210
Leu	Asn	Arg	Gly	Val	Ile	Lys	Gln	Val	Lys	Gly	Lys	Gly	Ala	Ser
				215					220					225
Gly	Ser	Phe	Val	Val	Val	Gln	Lys	Ser	Arg	Lys	Thr	Pro	Gln	Lys
				230					235					240
Ser	Arg	Asn	Arg	Lys	Asn	Arg	Ser	Ser	Ala	Val	Asp	Pro	Glu	Pro
				245					250					255
Gln	Val	Lys	Leu	Glu	Asp	Val	Leu	Pro	Leu	Ala	Phe	Thr	Arg	Leu
				260					265					270
Cys	Glu	Pro	Lys	Glu	Ala	Ser	Tyr	Ser	Leu	Ile	Arg	Lys	Tyr	Val
				275					280					285
Ser	Gln	Tyr	Tyr	Pro	Lys	Leu	Arg	Val	Asp	Ile	Arg	Pro	Gln	Leu
				290					295					300
Leu	Lys	Asn	Ala	Leu	Gln	Arg	Ala	Val	Glu	Arg	Gly	Gln	Leu	Glu
				305					310					315
Gln	Ile	Thr	Gly	Lys	Gly	Ala	Ser	Gly	Thr	Phe	Gln	Leu	Lys	Lys
				320					325					330
Ser	Gly	Glu	Lys	Pro	Leu	Leu	Gly	Gly	Ser	Leu	Met	Glu	Tyr	Ala
				335					340					345
Ile	Leu	Ser	Ala	Ile	Ala	Ala	Met	Asn	Glu	Pro	Lys	Thr	Cys	Ser
				350					355					360
Thr	Thr	Ala	Leu	Lys	Lys	Tyr	Val	Leu	Glu	Asn	His	Pro	Gly	Thr
				365					370					375
Asn	Ser	Asn	Tyr	Gln	Met	His	Leu	Leu	Lys	Lys	Thr	Leu	Gln	Lys
				380					385					390
Cys	Glu	Lys	Asn	Gly	Trp	Met	Glu	Gln	Ile	Ser	Gly	Lys	Gly	Phe
				395					400					405
Ser	Gly	Thr	Phe	Gln	Leu	Cys	Phe	Pro	Tyr	Tyr	Pro	Ser	Pro	Gly
				410					415					420
Val	Leu	Phe	Pro	Lys	Lys	Glu	Pro	Asp	Asp	Ser	Arg	Asp	Glu	Asp
				425					430					435
Glu	Asp	Glu	Asp	Glu	Ser	Ser	Glu	Glu	Asp	Ser	Glu	Asp	Glu	Glu
				440					445					450
Pro	Pro	Pro	Lys	Arg	Arg	Leu	Gln	Lys	Lys	Thr	Pro	Ala	Lys	Ser
				455					460					465
Pro	Gly	Lys	Ala	Ala	Ser	Val	Lys	Gln	Arg	Gly	Ser	Lys	Pro	Ala
				470					475					480
Pro	Lys	Val	Ser	Ala	Ala	Gln	Arg	Gly	Lys	Ala	Arg	Pro	Leu	Pro
				485					490					495
Lys	Lys	Ala	Pro	Pro	Lys	Ala	Lys	Thr	Pro	Ala	Lys	Lys	Thr	Arg
				500					505					510
Pro	Ser	Ser	Thr	Val	Ile	Lys	Lys	Pro	Ser	Gly	Gly	Ser	Ser	Lys
				515					520					525
Lys	Pro	Ala	Thr	Ser	Ala	Arg	Lys	Glu	Val	Lys	Leu	Pro	Gly	Lys
				530					535					540
Gly	Lys	Ser	Thr	Met	Lys	Lys	Ser	Phe	Arg	Val	Lys	Lys		
				545					550					

<210> 24
 <211> 461
 <212> PRT
 <213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2939607CD1

<400> 24

Met	Asp	Ala	Lys	Ser	Leu	Thr	Ala	Trp	Ser	Arg	Thr	Leu	Val	Thr	
1				5					10					15	
Phe	Lys	Asp	Val	Phe	Val	Asp	Phe	Thr	Arg	Glu	Glu	Trp	Lys	Leu	
				20					25					30	
Leu	Asp	Thr	Ala	Gln	Gln	Ile	Val	Tyr	Arg	Asn	Val	Met	Leu	Glu	
				35					40					45	
Asn	Tyr	Lys	Asn	Leu	Val	Ser	Leu	Gly	Tyr	Gln	Leu	Thr	Lys	Pro	
				50					55					60	
Asp	Val	Ile	Leu	Arg	Leu	Glu	Lys	Gly	Glu	Glu	Pro	Trp	Leu	Val	
				65					70					75	
Glu	Arg	Glu	Ile	His	Gln	Glu	Thr	His	Pro	Asp	Ser	Glu	Thr	Ala	
				80					85					90	
Phe	Glu	Ile	Lys	Ser	Ser	Val	Ser	Ser	Arg	Ser	Ile	Phe	Lys	Asp	
				95					100					105	
Lys	Gln	Ser	Cys	Asp	Ile	Lys	Met	Glu	Gly	Met	Ala	Arg	Asn	Asp	
				110					115					120	
Leu	Trp	Tyr	Leu	Ser	Leu	Glu	Glu	Val	Trp	Lys	Cys	Arg	Asp	Gln	
				125					130					135	
Leu	Asp	Lys	Tyr	Gln	Glu	Asn	Pro	Glu	Arg	His	Leu	Arg	Gln	Val	
				140					145					150	
Ala	Phe	Thr	Gln	Lys	Lys	Val	Leu	Thr	Gln	Glu	Arg	Val	Ser	Glu	
				155					160					165	
Ser	Gly	Lys	Tyr	Gly	Gly	Asn	Cys	Leu	Leu	Pro	Ala	Gln	Leu	Val	
				170					175					180	
Leu	Arg	Glu	Tyr	Phe	His	Lys	Arg	Asp	Ser	His	Thr	Lys	Ser	Leu	
				185					190					195	
Lys	His	Asp	Leu	Val	Leu	Asn	Gly	His	Gln	Asp	Ser	Cys	Ala	Ser	
				200					205					210	
Asn	Ser	Asn	Glu	Cys	Gly	Gln	Thr	Phe	Cys	Gln	Asn	Ile	His	Leu	
				215					220					225	
Ile	Gln	Phe	Ala	Arg	Thr	His	Thr	Gly	Asp	Lys	Ser	Tyr	Lys	Cys	
				230					235					240	
Pro	Asp	Asn	Asp	Asn	Ser	Leu	Thr	His	Gly	Ser	Ser	Leu	Gly	Ile	
				245					250					255	
Ser	Lys	Gly	Ile	His	Arg	Glu	Lys	Pro	Tyr	Glu	Cys	Lys	Glu	Cys	
				260					265					270	
Gly	Lys	Phe	Phe	Ser	Trp	Arg	Ser	Asn	Leu	Thr	Arg	His	Gln	Leu	
				275					280					285	
Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys	Lys	Glu	Cys	Gly	Lys	
				290					295					300	
Ser	Tyr	Ser	Gln	Arg	Ser	His	Leu	Val	Val	His	His	Arg	Ile	His	
				305					310					315	
Thr	Gly	Leu	Lys	Pro	Phe	Glu	Cys	Lys	Asp	Cys	Gly	Lys	Cys	Phe	
				320					325					330	
Ser	Arg	Ser	Ser	His	Leu	Tyr	Ser	His	Gln	Arg	Thr	His	Thr	Gly	
				335					340					345	
Glu	Lys	Pro	Tyr	Glu	Cys	His	Asp	Cys	Gly	Lys	Ser	Phe	Ser	Gln	
				350					355					360	
Ser	Ser	Ala	Leu	Ile	Val	His	Gln	Arg	Ile	His	Thr	Gly	Glu	Lys	
				365					370					375	
Pro	Tyr	Glu	Cys	Cys	Gln	Cys	Gly	Lys	Ala	Phe	Ile	Arg	Lys	Asn	
				380					385					390	
Asp	Leu	Ile	Lys	His	Gln	Arg	Ile	His	Val	Gly	Glu	Glu	Thr	Tyr	
				395					400					405	
Lys	Cys	Asn	Gln	Cys	Gly	Ile	Ile	Phe	Ser	Gln	Asn	Ser	Pro	Phe	
				410					415					420	
Ile	Val	His	Gln	Ile	Ala	His	Thr	Gly	Glu	Gln	Phe	Leu	Thr	Cys	
				425					430					435	
Asn	Gln	Cys	Gly	Thr	Ala	Leu	Val	Asn	Thr	Ser	Asn	Leu	Ile	Gly	
				440					445					450	
Tyr	Gln	Thr	Asn	His	Ile	Arg	Glu	Lys	Ala	Tyr					
				455					460						

<210> 25
 <211> 159
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3098421CD1

<400> 25
 Met Asp Lys Pro Arg Lys Glu Asn Glu Glu Glu Pro Gln Ser Ala
 1 5 10 15
 Pro Lys Thr Asp Glu Glu Arg Pro Pro Val Glu His Ser Pro Glu
 20 25 30
 Lys Gln Ser Pro Glu Glu Gln Ser Ser Glu Glu Gln Ser Ser Glu
 35 40 45
 Glu Glu Phe Phe Pro Glu Glu Leu Leu Pro Glu Leu Leu Pro Glu
 50 55 60
 Met Leu Leu Ser Glu Glu Arg Pro Pro Gln Glu Gly Leu Ser Arg
 65 70 75
 Lys Asp Leu Phe Glu Gly Arg Pro Pro Met Glu Gln Pro Pro Cys
 80 85 90
 Gly Val Gly Lys His Lys Leu Glu Glu Gly Ser Phe Lys Glu Arg
 95 100 105
 Leu Ala Arg Ser Arg Pro Gln Phe Arg Gly Asp Ile His Gly Arg
 110 115 120
 Asn Leu Ser Asn Glu Glu Met Ile Gln Ala Ala Asp Glu Leu Glu
 125 130 135
 Glu Met Lys Arg Val Arg Asn Lys Leu Met Ile Met His Trp Lys
 140 145 150
 Ala Lys Arg Ser Arg Pro Tyr Pro Ile
 155

<210> 26
 <211> 373
 <212> PRT
 <213> Homo sapiens

<220>
 <221> unsure
 <222> 368
 <223> unknown, or other

<220>
 <221> misc_feature
 <223> Incyte ID No: 3296650CD1

<400> 26
 Met Lys Ala Leu Phe Lys His Glu Ser Leu Gly Ser Gln Pro Leu
 1 5 10 15
 His Asp Arg Val Leu Gln Val Pro Gly Leu Ala Gln Gly Gly Cys
 20 25 30
 Cys Arg Glu Asp Ala Met Val Ala Ser Arg Leu Thr Pro Gly Ser
 35 40 45
 Gln Gly Leu Leu Lys Met Glu Asp Val Ala Leu Thr Leu Thr Pro
 50 55 60
 Gly Trp Thr Gln Leu Asp Ser Ser Gln Val Asn Leu Tyr Arg Asp
 65 70 75
 Glu Lys Gln Glu Asn His Ser Ser Leu Val Ser Leu Gly Gly Glu
 80 85 90
 Ile Gln Thr Lys Ser Arg Asp Leu Pro Pro Val Lys Lys Leu Pro
 95 100 105
 Glu Lys Glu His Gly Lys Ile Cys His Leu Arg Glu Asp Ile Ala
 110 115 120
 Gln Ile Pro Thr His Ala Glu Ala Gly Glu Gln Glu Gly Arg Leu
 125 130 135
 Gln Arg Lys Gln Lys Asn Ala Ile Gly Ser Arg Arg His Tyr Cys

His Glu Cys Gly	140	Lys Ser Phe Ala Gln	145	Ser Ser Gly Leu Thr	150
His Arg Arg Ile	155	His Thr Gly Glu Lys	160	Pro Tyr Glu Cys Glu	165
Cys Gly Lys Thr	170	Phe Ile Gly Ser Ser	175	Ala Leu Val Ile His	180
Arg Val His Thr	185	Gly Glu Lys Pro Tyr	190	Glu Cys Glu Glu Cys	195
Lys Val Phe Ser	200	His Ser Ser Asn Leu	205	Ile Lys His Gln Arg	210
His Thr Gly Glu	215	Lys Pro Tyr Glu Cys	220	Asp Asp Cys Gly Lys	225
Phe Ser Gln Ser	230	Cys Ser Leu Leu Glu	235	His His Lys Ile His	240
Gly Glu Lys Pro	245	Tyr Gln Cys Asn Met	250	Cys Gly Lys Ala Phe	255
Arg Asn Ser His	260	Leu Leu Arg His Gln	265	Arg Ile His Gly Asp	270
Asn Val Gln Asn	275	Pro Glu His Gly Glu	280	Ser Trp Glu Ser Gln	285
Arg Thr Glu Ser	290	Gln Trp Glu Asn Thr	295	Glu Ala Pro Val Ser	300
Lys Cys Asn Glu	305	Cys Glu Arg Ser Phe	310	Thr Arg Asn Arg Ser	315
Ile Glu His Gln	320	Lys Ile His Thr Gly	325	Asp Lys Pro Tyr Gln	330
Asp Thr Cys Gly	335	Lys Gly Phe Thr Arg	340	Thr Ser Tyr Leu Val	345
His Gln Arg Ser	350	His Val Gly Xaa Lys	355	Thr Leu Ser Gln	360
	365		370		

<210> 27

<211> 330

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3687719CD1

<400> 27

Met Gln Gln Gln	Pro	Leu	Pro	Gly	Pro	Gly	Ala	Pro	Thr	Thr	Glu
1	5					10					15
Pro Thr Lys Pro	Pro	Tyr	Ser	Tyr	Ile	Ala	Leu	Ile	Ala	Met	Ala
	20					25					30
Ile Gln Ser Ser	Pro	Gly	Gln	Arg	Ala	Thr	Leu	Ser	Gly	Ile	Tyr
	35					40					45
Arg Tyr Ile Met	Gly	Arg	Phe	Ala	Phe	Tyr	Arg	His	Asn	Arg	Pro
	50					55					60
Gly Trp Gln Asn	Ser	Ile	Arg	His	Asn	Leu	Ser	Leu	Asn	Glu	Cys
	65					70					75
Phe Val Lys Val	Pro	Arg	Asp	Asp	Arg	Lys	Pro	Gly	Lys	Gly	Ser
	80					85					90
Tyr Trp Thr Leu	Asp	Pro	Asp	Cys	His	Asp	Met	Phe	Glu	His	Gly
	95					100					105
Ser Phe Leu Arg	Arg	Arg	Arg	Arg	Phe	Thr	Arg	Gln	Thr	Gly	Ala
	110					115					120
Glu Gly Thr Arg	Gly	Pro	Ala	Lys	Ala	Arg	Arg	Gly	Pro	Leu	Arg
	125					130					135
Ala Thr Ser Gln	Asp	Pro	Gly	Val	Pro	Asn	Ala	Thr	Thr	Gly	Arg
	140					145					150
Gln Cys Ser Phe	Pro	Pro	Glu	Leu	Pro	Asp	Pro	Lys	Gly	Leu	Ser
	155					160					165
Phe Gly Gly Leu	Val	Gly	Ala	Met	Pro	Ala	Ser	Met	Cys	Pro	Ala
	170					175					180
Thr Thr Asp Gly	Arg	Pro	Arg	Pro	Pro	Met	Glu	Pro	Lys	Glu	Ile

Ser Thr Pro Lys	185	Pro Ala Cys Pro Gly	190	Glu Leu Pro Val Ala Thr	195
Ser Ser Ser Ser	200	Cys Pro Ala Phe Gly	205	Pro Ala Gly Phe Ser	210
Glu Ala Glu Ser	215	Phe Asn Lys Ala Pro	220	Thr Pro Val Leu Ser	225
Glu Ser Gly Ile	230	Gly Ser Ser Tyr Gln	235	Arg Leu Gln Ala Leu	240
Asn Phe Cys Met	245	Gly Ala Asp Pro Gly	250	Leu Glu His Leu Leu	255
Ser Ala Ala Pro	260	Ser Pro Ala Pro Pro	265	Thr Pro Pro Gly Ser	270
Arg Ala Pro Leu	275	Pro Leu Pro Thr Asp	280	Lys Glu Pro Trp Val	285
Ala Gly Gly Phe	290	Pro Val Gln Gly Gly	295	Ser Gly Tyr Pro Leu	300
Leu Thr Pro Cys	305	Leu Tyr Arg Thr Pro	310	Gly Met Phe Phe Phe	315
	320		325		330

<210> 28

<211> 396

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3774188CD1

<400> 28

Met Lys Ala Val Lys	5	Ser Glu Arg Glu Arg	10	Gly Ser Arg Arg Arg	15
His Arg Asp Gly Asp	20	Val Val Leu Pro Ala	25	Gly Val Val Val Lys	30
Gln Glu Arg Leu Ser	35	Pro Glu Val Ala Pro	40	Pro Ala His Arg Arg	45
Pro Asp His Ser Gly	50	Gly Ser Pro Ser Pro	55	Pro Thr Ser Glu Pro	60
Ala Arg Ser Gly His	65	Arg Gly Asn Arg Ala	70	Arg Gly Val Ser Arg	75
Ser Pro Pro Lys Lys	80	Lys Asn Lys Ala Ser	85	Gly Arg Arg Ser Lys	90
Ser Pro Arg Ser Lys	95	Arg Asn Arg Ser Pro	100	His His Ser Thr Val	105
Lys Val Lys Gln Glu	110	Arg Glu Asp His Pro	115	Arg Arg Gly Arg Glu	120
Asp Arg Gln His Arg	125	Glu Pro Ser Glu Gln	130	Glu His Arg Arg Ala	135
Arg Asn Ser Asp Arg	140	Asp Arg His Arg Gly	145	His Ser His Gln Arg	150
Arg Thr Ser Asn Glu	155	Arg Pro Gly Ser Gly	160	Gln Gly Gln Gly Arg	165
Asp Arg Asp Thr Gln	170	Asn Leu Gln Ala Gln	175	Glu Glu Glu Arg Glu	180
Phe Tyr Asn Ala Arg	185	Arg Arg Glu His Arg	190	Gln Arg Asn Asp Val	195
Gly Gly Gly Gly Ser	200	Glu Ser Gln Glu Leu	205	Val Pro Arg Pro Gly	210
Gly Asn Asn Lys Glu	215	Lys Glu Val Pro Ala	220	Lys Glu Lys Pro Ser	225
Phe Glu Leu Ser Gly	230	Ala Leu Leu Glu Asp	235	Thr Asn Thr Phe Arg	240
Gly Val Val Ile Lys	245	Tyr Ser Glu Pro Pro	250	Glu Ala Arg Ile Pro	255
Lys Lys Arg Trp Arg	260	Leu Tyr Pro Phe Lys	265	Asn Asp Glu Val Leu	270
Pro Val Met Tyr Ile		His Arg Gln Ser Ala		Tyr Leu Leu Gly Arg	

His Arg Arg Ile	275	Ala Asp Ile Pro Ile	280	Asp His Pro Ser Cys	285
Lys Gln His Ala	290	Val Phe Gln Tyr Arg	295	Val Glu Tyr Thr	300
Ala Asp Gly Thr	305	Val Gly Arg Arg Val	310	Lys Pro Tyr Ile Ile	315
Leu Gly Ser Gly	320	Asn Gly Thr Phe Leu	325	Asn Asn Lys Arg Ile	330
Pro Gln Arg Tyr	335	Tyr Glu Leu Lys Glu	340	Lys Asp Val Leu Lys	345
Gly Phe Ser Ser	350	Arg Glu Tyr Val Leu	355	Leu His Glu Ser Ser	360
Thr Ser Glu Ile	365	Asp Arg Lys Asp Asp	370	Glu Asp Glu Glu Glu	375
Glu Glu Val Ser	380	Asp Ser	385		390
	395				

<210> 29
 <211> 126
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4349106CD1

Met Gly Leu Leu Thr	1	Phe Arg Asp Val	5	Ala Ile Glu Phe Ser Arg	15
Glu Glu Trp Glu His	20	Leu Asp Ser Asp	25	Gln Lys Leu Leu Tyr Gly	30
Asp Val Met Leu Glu	35	Asn Tyr Gly Asn Leu	40	Val Ser Leu Gly Leu	45
Ala Val Ser Lys Pro	50	Asp Leu Ile Thr Phe	55	Leu Glu Gln Arg Lys	60
Glu Pro Trp Asn Val	65	Lys Ser Ala Glu Thr	70	Val Ala Ile Gln Pro	75
Asp Ile Phe Ser His	80	Asp Thr Gln Gly Leu	85	Leu Arg Lys Lys Leu	90
Ile Glu Ala Ser Phe	95	Gln Lys Val Ile Leu	100	Asp Gly Tyr Gly Ser	105
Cys Gly Pro Gln Asn	110	Leu Asn Leu Arg Lys	115	Glu Trp Glu Ser Glu	120
Gly Lys Ile Ile Leu	125	Trp			

<210> 30
 <211> 519
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4834217CD1

Met Ala Ala Glu Ala	1	Ala Asp Leu Gly Leu	10	Gly Ala Ala Val Pro	15
Val Glu Leu Arg Arg	20	Glu Arg Arg Met Val	25	Cys Val Glu Tyr Pro	30
Gly Val Val Arg Asp	35	Val Ala Lys Met Leu	40	Pro Thr Leu Gly Gly	45
Glu Glu Gly Val Ser	50	Arg Ile Tyr Ala Asp	55	Pro Thr Lys Arg Leu	60
Glu Leu Tyr Phe Arg	65	Pro Lys Asp Pro Tyr	70	Cys His Pro Val Cys	75

Ala	Asn	Arg	Phe	Ser	Thr	Ser	Ser	Leu	Leu	Leu	Arg	Ile	Arg	Lys
				80					85					90
Arg	Thr	Arg	Arg	Gln	Lys	Gly	Val	Leu	Gly	Thr	Glu	Ala	His	Ser
				95					100					105
Glu	Val	Thr	Phe	Asp	Met	Glu	Ile	Leu	Gly	Ile	Ile	Ser	Thr	Ile
				110					115					120
Tyr	Lys	Phe	Gln	Gly	Met	Ser	Asp	Phe	Gln	Tyr	Leu	Ala	Val	His
				125					130					135
Thr	Glu	Ala	Gly	Gly	Lys	His	Thr	Ser	Met	Tyr	Asp	Lys	Val	Leu
				140					145					150
Met	Leu	Arg	Pro	Glu	Lys	Glu	Ala	Phe	Phe	His	Gln	Glu	Leu	Pro
				155					160					165
Leu	Tyr	Ile	Pro	Pro	Pro	Ile	Phe	Ser	Arg	Leu	Asp	Ala	Pro	Val
				170					175					180
Asp	Tyr	Phe	Tyr	Arg	Pro	Glu	Thr	Gln	His	Arg	Glu	Gly	Tyr	Asn
				185					190					195
Asn	Pro	Pro	Ile	Ser	Gly	Glu	Asn	Leu	Ile	Gly	Leu	Ser	Arg	Ala
				200					205					210
Arg	Arg	Pro	His	Asn	Ala	Ile	Phe	Val	Asn	Phe	Glu	Asp	Glu	Glu
				215					220					225
Val	Pro	Lys	Gln	Pro	Leu	Glu	Ala	Ala	Ala	Gln	Thr	Trp	Arg	Arg
				230					235					240
Val	Cys	Thr	Asn	Pro	Val	Asp	Arg	Lys	Val	Glu	Glu	Glu	Leu	Arg
				245					250					255
Lys	Leu	Phe	Asp	Ile	Arg	Pro	Ile	Trp	Ser	Arg	Asn	Ala	Val	Lys
				260					265					270
Ala	Asn	Ile	Ser	Val	His	Pro	Asp	Lys	Leu	Lys	Val	Leu	Leu	Pro
				275					280					285
Phe	Ile	Ala	Tyr	Tyr	Met	Ile	Thr	Gly	Pro	Trp	Arg	Ser	Leu	Trp
				290					295					300
Ile	Arg	Phe	Gly	Tyr	Asp	Pro	Arg	Lys	Asn	Pro	Asp	Ala	Lys	Ile
				305					310					315
Tyr	Gln	Val	Leu	Asp	Phe	Arg	Ile	Arg	Cys	Gly	Met	Lys	His	Gly
				320					325					330
Tyr	Ala	Pro	Ser	Asp	Leu	Pro	Val	Lys	Ala	Lys	Arg	Ser	Thr	Tyr
				335					340					345
Asn	Tyr	Ser	Leu	Pro	Ile	Thr	Val	Lys	Lys	Thr	Ser	Ser	Gln	Leu
				350					355					360
Val	Thr	Met	His	Asp	Leu	Lys	Gln	Gly	Leu	Gly	Pro	Ser	Gly	Thr
				365					370					375
Ser	Gly	Ala	Arg	Lys	Pro	Ala	Ser	Ser	Lys	Tyr	Lys	Leu	Lys	Asp
				380					385					390
Ser	Val	Tyr	Ile	Phe	Arg	Glu	Gly	Ala	Leu	Pro	Pro	Tyr	Arg	Gln
				395					400					405
Met	Phe	Tyr	Gln	Leu	Cys	Asp	Leu	Asn	Val	Glu	Glu	Leu	Gln	Lys
				410					415					420
Ile	Ile	His	Arg	Asn	Asp	Gly	Ala	Glu	Asn	Ser	Cys	Thr	Glu	Arg
				425					430					435
Asp	Gly	Trp	Cys	Leu	Pro	Lys	Thr	Ser	Asp	Glu	Leu	Arg	Asp	Thr
				440					445					450
Met	Ser	Leu	Met	Ile	Arg	Gln	Thr	Ile	Arg	Ser	Lys	Arg	Pro	Ala
				455					460					465
Leu	Phe	Ser	Ser	Ser	Ala	Lys	Ala	Asp	Gly	Gly	Lys	Glu	Gln	Leu
				470					475					480
Thr	Tyr	Glu	Ser	Gly	Glu	Asp	Glu	Glu	Asp	Glu	Glu	Glu	Glu	Glu
				485					490					495
Glu	Glu	Glu	Glu	Asp	Phe	Lys	Pro	Ser	Asp	Gly	Ser	Glu	Asn	Glu
				500					505					510
Met	Glu	Thr	Glu	Ile	Leu	Asp	Tyr	Val						
				515										

<210> 31
 <211> 493
 <212> PRT
 <213> Homo sapiens
 <220>

<221> misc_feature

<223> Incyte ID No: 5156094CD1

<400> 31

```

Met Ile Leu Asn Ser Leu Ser Leu Cys Tyr His Asn Lys Leu Ile
 1      5      10      15
Leu Ala Pro Met Val Arg Val Gly Thr Leu Pro Met Arg Leu Leu
 20      25      30
Ala Leu Asp Tyr Gly Ala Asp Ile Val Tyr Cys Glu Glu Leu Ile
 35      40      45
Asp Leu Lys Met Ile Gln Cys Lys Arg Val Val Asn Glu Val Leu
 50      55      60
Ser Thr Val Asp Phe Val Ala Pro Asp Asp Arg Val Val Phe Arg
 65      70      75
Thr Cys Glu Arg Glu Gln Asn Arg Val Val Phe Gln Met Gly Thr
 80      85      90
Ser Asp Ala Glu Arg Ala Leu Ala Val Ala Arg Leu Val Glu Asn
 95      100     105
Asp Val Ala Gly Ile Asp Val Asn Met Gly Cys Pro Lys Gln Tyr
110     115     120
Ser Thr Lys Gly Gly Met Gly Ala Ala Leu Leu Ser Asp Pro Asp
125     130     135
Lys Ile Glu Lys Ile Leu Ser Thr Leu Val Lys Gly Thr Arg Arg
140     145     150
Pro Val Thr Cys Lys Ile Arg Ile Leu Pro Ser Leu Glu Asp Thr
155     160     165
Leu Ser Leu Val Lys Arg Ile Glu Arg Thr Gly Ile Ala Ala Ile
170     175     180
Ala Val His Gly Arg Lys Arg Glu Glu Arg Pro Gln His Pro Val
185     190     195
Ser Cys Glu Val Ile Lys Ala Ile Ala Asp Thr Leu Ser Ile Pro
200     205     210
Val Ile Ala Asn Gly Gly Ser His Asp His Ile Gln Gln Tyr Ser
215     220     225
Asp Ile Glu Asp Phe Arg Gln Ala Thr Ala Ala Ser Ser Val Met
230     235     240
Val Ala Arg Ala Ala Met Trp Asn Pro Ser Ile Phe Leu Lys Glu
245     250     255
Gly Leu Arg Pro Leu Glu Glu Val Met Gln Lys Tyr Ile Arg Tyr
260     265     270
Ala Val Gln Tyr Asp Asn His Tyr Thr Asn Thr Lys Tyr Cys Leu
275     280     285
Cys Gln Met Leu Arg Glu Gln Leu Glu Ser Pro Gln Gly Arg Leu
290     295     300
Leu His Ala Ala Gln Ser Ser Arg Glu Ile Cys Glu Ala Phe Gly
305     310     315
Leu Gly Ala Phe Tyr Glu Glu Thr Thr Gln Glu Leu Asp Ala Gln
320     325     330
Gln Ala Arg Leu Ser Ala Lys Thr Ser Glu Gln Thr Gly Glu Pro
335     340     345
Ala Glu Asp Thr Ser Gly Val Ile Lys Met Ala Val Lys Phe Asp
350     355     360
Arg Arg Ala Tyr Pro Ala Gln Ile Thr Pro Lys Met Cys Leu Leu
365     370     375
Glu Trp Cys Arg Arg Glu Lys Leu Ala Gln Pro Val Tyr Glu Thr
380     385     390
Val Gln Arg Pro Leu Asp Arg Leu Phe Ser Ser Ile Val Thr Val
395     400     405
Ala Glu Gln Lys Tyr Gln Ser Thr Leu Trp Asp Lys Ser Lys Lys
410     415     420
Leu Ala Glu Gln Ala Ala Ala Ile Val Cys Leu Arg Ser Gln Gly
425     430     435
Leu Pro Glu Gly Arg Leu Gly Glu Glu Ser Pro Ser Leu His Lys
440     445     450
Arg Lys Arg Glu Ala Pro Asp Gln Asp Pro Gly Gly Pro Arg Ala
455     460     465
Gln Glu Leu Ala Gln Pro Gly Asp Leu Cys Lys Lys Pro Phe Val

```

Ala Leu Gly Ser 470 Gly Glu Glu Ser Pro 475 Leu Glu Gly Trp 480
 485 490

<210> 32
 <211> 516
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5665139CD1

<400> 32
 Met His Gly Arg Lys Asp Asp Ala Gln Lys Gln Pro Val Lys Asn
 1 5 10 15
 Gln Leu Gly Leu Asn Pro Gln Ser His Leu Pro Glu Leu Gln Leu
 20 25 30
 Phe Gln Ala Glu Gly Lys Ile Tyr Lys Tyr Asp His Met Glu Lys
 35 40 45
 Ser Val Asn Ser Ser Ser Leu Val Ser Pro Pro Gln Arg Ile Ser
 50 55 60
 Ser Thr Val Lys Thr His Ile Ser His Thr Tyr Glu Cys Asn Phe
 65 70 75
 Val Asp Ser Leu Phe Thr Gln Lys Glu Lys Ala Asn Ile Gly Thr
 80 85 90
 Glu His Tyr Lys Cys Asn Glu Arg Gly Lys Ala Phe His Gln Gly
 95 100 105
 Leu His Phe Thr Ile His Gln Ile Ile His Thr Lys Glu Thr Gln
 110 115 120
 Phe Lys Cys Asp Ile Cys Gly Lys Ile Phe Asn Lys Lys Ser Asn
 125 130 135
 Leu Ala Ser His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Lys
 140 145 150
 Cys Asn Glu Cys Gly Lys Val Phe His Asn Met Ser His Leu Ala
 155 160 165
 Gln His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Asn
 170 175 180
 Glu Cys Gly Lys Val Phe Asn Gln Ile Ser His Leu Ala Gln His
 185 190 195
 Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Asn Glu Cys
 200 205 210
 Gly Lys Val Phe His Gln Ile Ser His Leu Ala Gln His Arg Thr
 215 220 225
 Ile His Thr Gly Glu Lys Pro Tyr Glu Cys Asn Lys Cys Gly Lys
 230 235 240
 Val Phe Ser Arg Asn Ser Tyr Leu Val Gln His Leu Ile Ile His
 245 250 255
 Thr Gly Glu Lys Pro Tyr Arg Cys Asn Val Cys Gly Lys Val Phe
 260 265 270
 His His Ile Ser His Leu Ala Gln His Gln Arg Ile His Thr Gly
 275 280 285
 Glu Lys Pro Tyr Lys Cys Asn Glu Cys Gly Lys Val Phe Ser His
 290 295 300
 Lys Ser Ser Leu Val Asn His Trp Arg Ile His Thr Gly Glu Lys
 305 310 315
 Pro Tyr Lys Cys Asn Glu Cys Gly Lys Val Phe Ser His Lys Ser
 320 325 330
 Ser Leu Val Asn His Trp Arg Ile His Thr Gly Glu Lys Pro Tyr
 335 340 345
 Lys Cys Asn Glu Cys Gly Lys Val Phe Ser Arg Asn Ser Tyr Lys
 350 355 360
 Ala Gln His Leu Ile Ile His Ala Gly Glu Lys Pro Tyr Lys Cys
 365 370 375
 Asp Glu Cys Asp Lys Ala Phe Ser Gln Asn Ser His Leu Val Gln
 380 385 390
 His His Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Asp Glu

				395					400					405
Cys	Gly	Lys	Val	Phe	Ser	Gln	Asn	Ser	Tyr	Leu	Ala	Tyr	His	Trp
				410					415					420
Arg	Ile	His	Thr	Gly	Glu	Lys	Ala	Tyr	Lys	Cys	Asn	Glu	Cys	Gly
				425					430					435
Lys	Val	Phe	Gly	Leu	Asn	Ser	Ser	Leu	Ala	His	His	Arg	Lys	Ile
				440					445					450
His	Thr	Gly	Glu	Lys	Pro	Phe	Lys	Cys	Asn	Glu	Cys	Gly	Lys	Ala
				455					460					465
Phe	Ser	Met	Arg	Ser	Ser	Leu	Thr	Asn	His	His	Ala	Ile	His	Thr
				470					475					480
Gly	Glu	Lys	His	Phe	Lys	Cys	Asn	Glu	Cys	Gly	Lys	Leu	Phe	Arg
				485					490					495
Asp	Asn	Ser	Tyr	Leu	Val	Arg	His	Gln	Arg	Phe	His	Ala	Gly	Lys
				500					505					510
Lys	Ser	Asn	Thr	Cys	Asn									
				515										

<210> 33

<211> 3163

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 091502CB1

<400> 33

```

gtcggagggtc ttacccaaca gattgacgcg gcgttagtat tggccgtgta cccgaaaaaac 60
tgattgactg ggctggcggtt aactgtgctg aggcaggatt tccctgggga aagtcttcac 120
tcagagcttc gctgacagcc aaggatggca agcaaggggc cctcggcctc tgcattctcct 180
gagaactcca gtgcaggggg gcccagtggt agcagcaatg gcgctggcga gagcggaggg 240
caggacagca ctttcgagt caacatctgc ttggacacag ccaaggatgc cgtcatcagc 300
ctgtgtggcc acctcttctg ttggccgtgt ttacatcagt ggttggagac cagacctaac 360
agacaggtgt gtctgttttg caaagctggc atcagccgag acaaggatcat cccctctctat 420
ggaaggggca gcactgggca acaggacccc agagagaaga cccctcctcg tcctcaagga 480
cagaggccag agccggagaa tagaggggga ttcaaggat ttggatttgg agatgggtggc 540
ttccagatgt cttttggaat tggggcattt ccttttggga tatttgccac agcatttaat 600
ataaatgatg ggcggcctcc tccagctgtc cctgggacac cccagtatgt ggacgagcag 660
ttcctgtcac gcctcttctt atttgtggcc ctggtgatca tgttctggct cctgattgcc 720
taatgctggg ctctgcctta catccgtggc agggctctgg actggtgacg tgccacccca 780
actctgggtg tttggcttcc tggctaactt tgactcctgg aatcagtggt atcagtaaca 840
catcaaggag tcttgtttct tcatcagagc tttggaactc gagaccagtt ggcgatgacc 900
cctgaatatc gccaccgtg taaacactct ataacttcag gccttggcat tgagtcatct 960
ctcatgggtg acaccatgaa atcttgtttc agccagttct gcaggtcctg actctgcaga 1020
gggaagaggc agaaagagag aaactgtcag agtataattt cacctgagtt taataattaca 1080
gaaacaaagg gatgcaccaa atggtatttc tggaaatttt catgtcttta aatacccctt 1140
ggtaagttag ttctgaagcc agtgggggct cctcagatag agaggttccc ctttcaaate 1200
ccagtgcgcg tctgttctct tctcttcccc tcccactccc cctcttcttc ctctgtagag 1260
atgcaagaaa ttgctgtccc ataaaaatca taattgcagt agctaaagct ggggtcactt 1320
cgtgaattca ccagagactc aaagatcttt tattggctct gggctgtgct cagtgtcttt 1380
ggcctcagag aacaacttga atgacttctt ggtttcctgg cataaattat tctgtgtgag 1440
acatgtggct taactcacag gtttcccatc agctttctcc ctaaaactat gttcatctgc 1500
ctctctctgc cagagaacat acagccgaga atactgccga agctgagact gactactgtg 1560
cattaggaaa gacctggagt caggactttg gtgggatttg gagctccgag gcagtaataa 1620
ctgaacaagc agccctgtcc ctaggctgac agaagcttga atgcactct tcccagaacc 1680
tgccacagga aactgggggc tttgtcaggt cagcccaact gcatgcaaaa gaccaccatc 1740
ctcagaagcc aagttgtctt ttatgaagag gcaaggaaaag gggaaaacca catgtgaccc 1800
tgatttttgt atggcttgat agagttccct gaaaactcct tgtatgtgtg ctaaaaccag 1860
ggaagcatgt gactgccaag caggcaaccc ctgattgatt gtaaagccag gtggcagggc 1920
cttggggcag cccagcaca tgatattgtg ttgtcttccc tctgtggaa tcgaggggaa 1980
attattcttc ccaatacctt gatttgattt tcagtttcat aagcttcttc ctctgaatct 2040
tattgaggga ctatggtacc aagcaggtag gactgttcac ctggtggaac agttcttgct 2100
ctgccttcta ggcttcaccc cagaaatcca gcctctttct ggagacccca aagctggagg 2160
gagatgggct ttctctctgg cctctctctt actttgccat ccactgtct cctgggtaac 2220
cccagcaata accaacaat ggtaggaagc cccatctatt gctttttctc aattatgact 2280
gcatagttta tggaaacaaa gatcttgagg aagatgaggg aagccctccc ctcttcacag 2340

```

```

tccccatctc ttctcctttg tacctgtcaa caccagagtt cagtgtttta acagacaaaa 2400
tataaagtat tgagtaggtg gttcatatgc cgaatccact tggtaggaag aaaccaccag 2460
ctttatagtg tgccctgatag attttaaaca ttccctgggca cccactcaag agtgctcttt 2520
ttatcacctt ctggaaatcc gcaaagttgc aggggcctct ggagtgtctc ttctctagag 2580
agaattgggt ggacccccct cagtgcagtg gcccaacta gtggaggagg agaggactta 2640
agtcagatgg actcaacaga aatgggtttc cagaagaata atgaaaagtt gtgggtagga 2700
aaatgaatca tttggactct tcaatgaaat ggagtgaagc caggagagct cagccaacag 2760
aggcactctg ggaacctgtt agtaaagcca ggctggccaa atgccatttg attttgaacc 2820
tcgtagggtc cactcaccct tctgccagga gctaagtaag gcaggagagc tgacttggga 2880
ctcctggctc ggcccccaaca gggagccccct ttcccaccat cctcgggcaa gctcaccacc 2940
tcacctcttc gccaaaggcag ctttccctttc ttttgtgtgt tttctgtgtt cttagcctcc 3000
accctcctcc tgccaccctt gtggactagg accaggtcct gaccccagtc agaaaatgat 3060
gatatgtaca gtggcacacc ttaaccagtc actaattttc actgttgtga aagtgat.ttg 3120
atttagaatt aaacaaatgg ttttacatta ctatgaaaaa aaa 3163

```

```

<210> 34
<211> 2185
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 763816CB1

```

```

<400> 34
ggatctccag agctcagaaa accctcaggg tcaccagatc tttggaagct ttctcctgat 60
cagcggaaaa cttctcctgc ttcacttgat ttccctgagt cccagaaaag ttcccggtgt 120
ggttctcctg atctctggaa gtcttccctt ttatttgagc ctcaaaaacc tgtcttccct 180
gagagccgaa aaccatgtcc ttatgggcca tctgaggccc caaaagcagc ctcatatata 240
tggaagctgt tctctctatc gatactgagc ctagaaaacc tgccctgttt cccgagcctg 300
ccaaaacagc cctcctctgt tctccagaag cagcgaacag tgcccttttt ccagagcccc 360
ggaagcatgc ccttttccct gaactcccca aatctgctct attctcagaa tcacagaagg 420
ctgttgagct tgggtgatga ctacaaatag atgccataga tgatcaaaaa tgtgatattt 480
tggttcagga agaacttcta gcttcaccta agaaactctt agaagatact ttatttccct 540
cctcaaagaa gctcaagaaa gacaaccaag agagctcaga cgctgagctt agtagtagtg 600
agtacataaa aacagatttg gatgcatggg atattaaggg ccaggaatca agcagtgatc 660
aagagcaggt tgatgtggaa tccattgatt ttagcaaaga gaacaaaatg gacatgacta 720
gtccagagca gtctagaat gtgctacagt ttactgaaga aaaagaagct tttatctctg 780
aagaggagat tgcaaaatc atgaagcgtg gaaaaggaaa gtattattgc aaaatttgtt 840
gctgtcgtgc tatgaaaaaa ggtgctgttt tgcattcatt ggttaataag cataatgttc 900
atagccctta caaatgcaca atctgtggaa aggcctttct tttggaatct ctcttaaaaa 960
atcatgtagc agcccatggg caaagtttac ttaaattgtc acgttgtaat tttgaatcaa 1020
atttcccaag aggttttaag aaacatttaa ctcatgttca aagccggcat aatgaaggag 1080
caataaaaaa gctaattgaa gctcttgaac cgccactgga ggagcagcaa atttgataac 1140
acagtgtgaa tatttgttct acaaagggtg ttgttggaac cattctttgt aagtatagct 1200
tatcagatag catagttgga tcagtagatg acatgtatgg tgtaccgtgt ttcactgtct 1260
cagtttgtgt actaagaatg agcatttgat ctttttttct tggctctctg ctatgtgact 1320
atcttgtaag tcaataaatt tctgtatagt ccagatggat taaacttctc atttctttta 1380
aatatgtatg aataataata caaggaagta ggcattccat ttaataatca agagcaagtt 1440
gtactcaaag cattcagtta aagtgtatct gtgtgtggaa ctaatttcag acaatagaaa 1500
atattagttg aaatgtttta gaattaggca tgaaaaataa atttgagaaa ttttgtttcc 1560
ttacatgtat ttttaaatca taagagtat tttctatctg atgtaaaatt agtttataaa 1620
tcttaatcag cttctagatg tttattagct tttatgtcat ttaataatca agagcaagtt 1680
ttgctgattt tctgctaatt ggaaaaattg actaagtctt taaaatagtt tgcagccttc 1740
tcccacagga gacaagtga agataagtg gatttttagat ctttctttgt catagttgtt 1800
ttcagtgagg tcttccattc tgtatcttac cctaagatct ggttcttccc tccccatccc 1860
cacccccac ccaccgctg ccagctcaca ctaatagatg attcttaatt gccaaatgtg 1920
ttagagtttg tatatcctac tctgggcct tacatgtcgc ctggtggggc ttaagaccag 1980
gttgataagt aggaactgaa agtcttccag attcacagta gaaaatttta tagacatttc 2040
tggttaaagaa atatatcgat tttatgtttt tcaattatgt tactgtaaat acctgtacc 2100
tgttcatgga ttattttatt ctaaaatatt ttgtcaaag tgtatcaacc aaattaaaaa 2160
gaaaggtttt catgtcaaaa aaaaa 2185

```

```

<210> 35
<211> 2408
<212> DNA
<213> Homo sapiens

```

<220>
<221> unsure
<222> 2243
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte ID No: 961184CB1

<400> 35
ggagcatccg gacgctccag tccaatctgg gctgcctccc accatcctcg gctttgacct 60
ctgggaccag gagcctccca cgccccattg acggtgtttc ggactggagc caaggggtgct 120
ccctgcgata cactggcagc cctgcctccc tggccagcaa cttggaaatc agccagtccc 180
ctaccatgcc ctctctcagc ctgcaccgca gccacatgg gccagcaag ctctgtgacg 240
acccccaggc cagcttggtg cccgagcctg tccccggtgg ctgccaggag cctgaggaga 300
tgagctggcc gccatcgggg gagattgcca gccaccaga gctgccaagc agccacctc 360
ctgggcttcc cgaagtggcc ccagatgcaa cctccactgg cctccctgat acccccgag 420
ctccagaaac cagcaccaac taccagtggt agtgcaccga ggggtctgca ggccccagct 480
ctctccctt gcctattctg gagccggtca aaaaccctg ctctgtcaaa gaccagacgc 540
cactccaact ttctgtagaa gataccacct ctccaaatac caagcctgac ccacctactc 600
ccaccacccc agaaacatcc cctcctctct cctcctctcc tcttctatct actccttgtt 660
cagctcacct gacccctctc tccctgttcc cttcctcctt ggaatcatca tcggaacaga 720
aattctataa ctttgtgatc ctccacgcca gggcagacga acacatcgcc ctgctgggttc 780
gggagaagct ggagggcctt ggctgtgccc acggggccac cttctgagag gatttccagg 840
tgccggggcg cggggagctg agctgcctgc aggcagccat agaccactca gctttcatca 900
tctacttct cacctccaac ttcgactgtc gcctgagcct gcaccaggtg aaccaagcca 960
tgatgagcaa cctcacgcga caggggtgc cagactgtgt catcccttc ctgcccctgg 1020
agagctcccc ggcccagctc agctccgaca cggccagcct gctctccggg ctggtgcggc 1080
tggacgaaca ctcccagatc ttccgaggga aggtggccaa caccttcaag cccacaggc 1140
ttcaggcccc aaaggccatg tggaggaagg aacaggacac ccgagccctg cgggaacaga 1200
gccaacacct ggacggtgag cggatgcagg cggcggcact gaacgcagcc tactcagcct 1260
acctccagag ctacttgtcc taccaggcac agatggagca gctccaggtg gcttttgga 1320
gccacatgtc atttgggact ggggcgcctt atggggctcg aatgccctt gggggccagg 1380
tgccccctgg agccccgcca ccttttccca cttggccggg gtgcccgcag ccgccacccc 1440
tgacgcgatg gcaggctggc accccccac cgccctcccc acagccagca gcctttccac 1500
agtcactgcc ctccccgcag tccccagcct tccctacggc ctacccgca cccctcaga 1560
gcccagggtc gcaaccctc attatccacc acgcacagat ggtacagctg gggctgaaca 1620
accacatgtg gaaccagaga gggtcccagg cgcccaggga caagacgcag gaggcagaat 1680
gaccgcgtgt ccttgccctga ccacctgggg aacaccctg gaccaggca tcggccagga 1740
ccccatagag caccocggtc tgccctgtgc cctgtggaca gtggaagatg aggtcatctg 1800
ccactttcag gacattgtcc gggagccctt catttaggac aaaacgggcg cgatgatgcc 1860
ctggctttca ggggtggtcag aactggatac ggtgtttaca attccaatct ctctatttct 1920
gggtgaaggg tcttggtggt gggggtattg ctacggctct ttaattataa taaatattta 1980
ttgaatgcta aaccatatca aacacttccc aaaatttaca agcaagagag agttaaaatt 2040
aggaatatata tctgtgtaaa taacaatctg cctatacatt gtcttataaa ttaatttctc 2100
tttacatagt tttgagtatt attttaagat tgcattcttt ccagtgttcc ttttgtcaaa 2160
aaatgttttc atgaggaagg agaattgtgt gtgtttacag agccaatgta tattaatttg 2220
agagaaactg atggggaagg ggnatcgact ttataggcag taaagtattt accaaccatg 2280
actgcctatt ttaaaaccgt aagccaaatt tgtattgtag ttaatccaaa ttggtcctgt 2340
taagtgaat tggttagaat ctgggtttta aaacttttta attttataaa ataaccaggg 2400
aggatttt 2408

<210> 36
<211> 2564
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1255525CB1

<400> 36
ccaacctcat tttttctctg ttgccacagg accctttgtg gaggagtgtt ttgaagtgac 60
atcctgttat ttccctatcg attttacccc tccacctaat gatccccatg gtatccagag 120
agaagacctc atcctgagtc ttgcgcgtgt gctggcttct acaccagat ttgctgagtt 180
tctgctgccc ctgttgattg agaaagtgga ttctgagggt ctgagtgcc agttggattc 240
tctacagact ctgaatgctt gctgtgctgt gtatggacag aaggaactga aggaattcct 300


```

ccccagcctt tgggcttcta tccgcagaga gaggagccag cggcggacaa tccttgaat 360
gctcctgggt ttcttgaagc tgcagcagaa atggagctat gaagacaaag atcaaaggcc 420
tctgaatggc ttcaaggacc agctgtgtctc atgggtctaa cagaccccag 480
caccagcctt cagcttgttg gcatccgtac actgacagtc ttgggtgccc agccagatct 540
cctatcttat gaggacttgg agctggcagt gggtcacctg tacagactga gcttccctgaa 600
ggaggattcc cagagttgca ggggtggcagc actggaagca tcaggaaccc tggctgtct 660
ctaccctgtg gccttcagca gccacctcgt acccaagctc gctgaggagc tgcgtgtagg 720
ggagtcaaat ttgactaacg gagatgagcc caccaatgc tcccggcatc tgtgtgtct 780
gcaagccttg tcagctgtat caacacatcc cagcatcgtc aaggagacac tgcctctgct 840
gctgcagcat ctctggcaag tgaacagagg gaatatgggt gcacaatcca gtgacgttat 900
tgtgtctgt cagagcctca gacagatggc agaaaaatgt cagcaggacc ctgagagttg 960
ctggtatttc caccagacag ctataccttg cctgcttgcc ttggctgtgc aggcctctat 1020
gccagagaag gagccctcag ttctgagaaa agtactattg gaggatgagg tgttggctgc 1080
catgggtgtct gtcattggca ctgctacaac ccacctgagc cctgagttag ctgcccagag 1140
tgtgacacac atttgtgccc tcttcttggg tggcaacgtg tcctttctgc ctgaaaacag 1200
cttcccagagc agattccagc cattccagga tggctctca gggcagaggc ggctgattgc 1260
actgcttatg gcctttgtct gctccctgcc tcgaaatgtg gaaatccctc agctgaacca 1320
actcatgcgg gagcttttgg aactgagctg ctgccacagc tgccctttt cttccaccgc 1380
tgctgccaa gcttttgag gactcctcaa caagcaccct gcagggcagc agctggatga 1440
attcctacag ctagctgtgg acaaagtggg ggctggcctg ggctctgggc cctgtcgtag 1500
tcaggccttc actcttcttc tctgggtaac aaaggcccta gtgctcagat accatcctct 1560
cagctcctgc cttacagccc ggctcatggg cctcctgagt gaccagaat taggtccagc 1620
agcagctgat ggcttctctc tgctcatgtc tgactgcact gatgtgctga ctcgtgctgg 1680
ccatgccgaa gtgaggatca tgttccgcca gcgttcttc acagataatg tgctgcttt 1740
ggccagggc ttccatgctg ctccccaaga tgtgaagcca aactacttga agggctcttc 1800
tcatgtactt aacaggctgc ccaagcctgt actcttgcca gagctgcca cgcttcttc 1860
cttgcctgctg gaggccctgt cctgccctga ctgtgtggtg cagctctcca cctcagctg 1920
ccttcagcct cttctactgg aagcacccca agtcatgagt cttcacgtgg acaccctcgt 1980
caccaagttt ctgaacctca gctctagccc ttccatggct gtccggatcg ccgcaactgca 2040
gtgcatgcat gctctcactc gcctgccac cctgtgctg ctgccgtaca aaccacaggt 2100
gattcggggc tttagcacaac ccctggatga caagaagaga ctggtgcgca aggaagcag 2160
gtcagccaga ggggagtggt ttctgttggg gacccctggc agctgagccc tgcactctg 2220
cctagactgt tctgacaatc taacctggga ttactaactg ttgagccatc ttcccaaag 2280
cagggaaaacc actggtctct gactgccttt cccacagaca cagcacaat gctaggcctc 2340
tggtgcatgg ctgtacaaag aacataagag tccatatctc tagtggattt gtaaaataag 2400
tgtgtgtgag acacttgctg ttgaagaaag atctagggtc ctgggtctct tgcattata 2460
tgtcagaaaa ggggcgatat gctgctgagg ggtgagtgca tatgagtgtg gccctgagga 2520
ccagggtggt cagatgttgt ctacctgctg aagaataaag atcg 2564

```

<210> 37
 <211> 957
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1297447CB1

```

<400> 37
ccttctgggc atgtctgcca tatggctcca ggtttgtttt tctccccggc actctgacgg 60
ggaggggtcc cggcatctcc tggcgctccg gtagaggacg cggaggatgc tgagctgctg 120
gcgcactgca gcacaactag agatgtacgg atgccccat cttgatctta cagaatcaga 180
ggtacagccg cgagaaagag tcaagaacag acagagtcgc ttgaggactc aggagggtgt 240
ttgctgcgtt gacaacagac tacaccctca cagtttgctc tgctcttcca acaccagtgg 300
aagatgatca catcccaggg atcagtgtcg tttaggatg tgactgtggg cttcactcaa 360
gaggagtggc agcatctgga cctgtctcag aggaccctgt acagggatgt gatgctggag 420
aactacagcc acctgtgtct agtaggggtat tgcattccta aaccagaagt gattctcaag 480
ttggagaaa gcgaggagcc atggatatta gaggaataat ttccaagcca gagtcatctg 540
ggtagttag tatgtgccag atggaattta aaggaaggta gatcacaag ggtaagtttg 600
gataataaga ccattgaaat gttctttagg aatcatgttt tagaggctcc agacctttg 660
aagtaacatg ttgacgtgac ccaaatatgc agtgactctg aatcaccag catctcccag 720
tatcactttc atacacacat cttgttgttt tttttaaat gtgatataaa ctatatagtt 780
tagccagaga ttaataaaat tttatatata tatgtaacac tcatgtaccc atcattgaga 840
tgaatatata aagtgtttt tcaacacccc cagaaagctc cctgttgtac tcttcccagt 900
aaagagcatc cctccttttg taaaattaaa tatatttttc tataaaaaaa aaaaaaa 957

```

<210> 38

<211> 2701
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1441094CB1

<400> 38
gcggcagtgta gtatgtgtgt gacggacccc cgcgcgccgc ggctcgggac cctgccttac 60
cctccttctt gccgcgccgg ctgtggaact agcgcgtgcc cctcgcggg cctcgacgtc 120
tcccgtccgc cctcactcg tggcagggcg cagctcctgc ttcaaggtta ctgactttt 180
atgatgtttg gtggctatga gactatagaa gcatacgaag atgatcttta tcgagatgag 240
tcacttagtg aactgagtg t gatagttag gtggaatttc aactctatag ccaaattcat 300
tatgcccaag atcttgatga tgtcatcagg gaggaagagc atgaagaaaa gaactctggg 360
aattcgggaat cttcgagtag taaaccaa atcagaagaagc taatcgtcct ttcagatagt 420
gaggtcatcc agctgtcaga tgggtcagag gtcatcactt tgtctgatga agacagtatt 480
tatagatgta aaggaaagaa tgttagagtt caagcacaag aaaatgccca tggctcttct 540
tcttctcttc aatctaata gctggttgat aagaaatgca agagtगतat tgagaagcct 600
aaatctgaag agagatcagg tgaatccga gaggtcatga ttatagaggt cagttcaagt 660
gaagaggaag agagcaccat ttcagaaggt gataatgtgg aaagctggat gctactggga 720
tgtgaagtag atgataaaga t gatgatct ctctcaacc ttgtgggatg tgaaaactct 780
gttactgaag gagaagatgg tataaactgg tccatcagtg acaaagacat tgaggccag 840
atagctaata accgaacacc tgggaagatgg acccagcgg actattcagc caacaaaaac 900
attatctgta gaaattgtga caaacgtgg catttatcaa aaaactgcc cttaccacga 960
aaagttcgtc gctgcttct gtgctccagg agaggacatc tctgtattc ctgtccagcc 1020
cccctttgtg aatactgtc tgtgcctaag atgtggacc actcatgtct tttcagacat 1080
tcctgggata aacagtgtga ccgagtcat atgctaggcc actatacaga tgcttgaca 1140
gaaatctgga ggcagtatca cctaacgacc aaacctggac caccacaaaa gccgaagacc 1200
ccttcaagac catcagcctt agcatattgc tatcactgcy cgcaaaaagg ccattatgga 1260
cacgaatgtc cagaagaga agtgtatgac ccgtctccag tatctccatt catctgtctac 1320
tatgatgaca aatatgaaat tcaggagaga gaaaagagac taaaacaaaa aataaaagta 1380
ctcaagaaaa atggggttat ccagagcca tccaagctac cttatataaa agcagcaaat 1440
gagaaccccc accatgatat aagggaagggc cgtgcctcat ggaaaagcaa caggtggcct 1500
caagaaaata aagaaacaca aaaagaaatg aagaacaaga atagaaactg ggagaagcac 1560
aggaaggctg acagacatcg tgaagtggat gaggattttc ccaggggcc caaacctac 1620
tcttctctg gcagttttaa aaccagaag ccttctaagc cctttcacg ttcacacat 1680
taccacacgt caagagaaga caagtctccc aaggaaggca agaggggcaa gcagaagaaa 1740
aaggagaggt gctgggaaga tgatgacaat gataacttat ttcttattaa gcagagaaaa 1800
aaaaagtctt aagccgtcag gcagcctctg atgtggcttt tcattgttca tttggccttt 1860
gtgtctataa ccttctggca ctgtgtttat tatctatgat taaataaagt gattttttgg 1920
ttttgttttt ttaatttcag ccattcctag agttactgaa tatccatgga gatctcaatt 1980
ctctgtgtcc aacaggatat taggtaagaa agtacaaaga taaacctgga ctctcctat 2040
tccaatatgt catctttact cagaatccta gggataggta gaagaattca tcttttcaag 2100
aaagtgtttt aaaaataact tgggaaaaaa actgcatcaa aggtaattta tctctaaatt 2160
aaatccttgc aggaagagaa attaacacta agaaaaagtc atcaatat ttcaactttt 2220
tttttttttt tttttacttt ggaaaggaca ataactataa acctatatcc agattttctt 2280
tctgtgtaag ctgtgtcag aatcttctt tggacaaaac atcactagct gactataaaa 2340
acaaaagtgt catcattgaa gccctgaaga ggcagggaat tgagcttcag caaaatacag 2400
gaaaagaact atccagtata aatgtcagaa gacagatttc ctaaacaagt aaaagagaca 2460
tcaaaaaatt taacataatc acaatgaaat cactttttac cacttttaca gcggtgtttc 2520
aagcggactg tctactcagat ctgcagagat gaatattact caaaaaattt tttgttctc 2580
ttgcattttt ttcaactacc gcaaagctat acagattttt ttgtacttgt ggatcttttg 2640
tactttctcat aacctaatgt cagactaaga aaaataaaat atctgagagt aaaaaaaaaa 2700
a 2701

<210> 39
<211> 2517
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1479382CB1

<400> 39
ggagaaaacg acttcagtgt catgtactcc acaaggaaaa actgtgctca actctggctg 60

```

ggctcctgctg cgtttataaaa ccatgattgc agacctaat gtaagtttgt gtcaactggg 120
cgagatacag catgtgtgaa ggctctaaga gacattgaac ctggagaaga aatttcttgt 180
tattatggag atgggttctt tggagaaaaat aatgagttct gcgagtgtta cacttgcgaa 240
agacggggca ctggtgcttt taaatccaga gtgggactgc ctgcgcctgc tcctgtttatc 300
aatagcaaat atggactcag agaaacagat aaacgtttta ataggcttaa aaagttagg 360
gacagcagca aaaattcaga cagtcaatct gtcagctcta acactgatgc agataccact 420
caggaaaaaa acaatgcaac ttctaaccga aaatcttcag ttggcgtaaa aaagaatagc 480
aagagcagaa cgttaacgag gcaatctatg tcaagaattc cagcttcttc caactctacc 540
tcatctaagc taactcatat aaataattcc aggggtacca agaaactgaa gaagcctgca 600
aagcctttac tttcaaagat aaaattgaga aatcattgca agcggctgga gcaaaagaat 660
gcttcaagaa aactcgaaat gggaaactta gtactgaaag agcctaaagt agttctgtat 720
aaaaatttgc ccattaaaaa agataaggag ccagagggag cagcccaagc cgcagttgcc 780
agcgggtgct tgactagaca cgcggcgaga gaacacagac agaatcctgt gagaggtgct 840
cattcgcagg gggagagctc gccctgcacc tacataactc ggcggctcagt gaggacaaga 900
acaaatctga aggaggcctc tgacatcaag cttgaacc aaacacagac agaatcctgt gagaggtgct 840
agcagtgtga cggaaacctt ccccgacagt ggtgaacagc tgcagccagc tcctgtgtctg 1020
caggaggaag aactggctca tgagactgca caaaaagggg aggcaaagtg tcataagagt 1080
gacacaggca tgtccaaaaa gaagtcacga caaggaaaac ttgtgaaaca gtttgcaaaa 1140
atagaggaat ctactccagt gcacgattct cctggaaaag acgacgcggt accagatttg 1200
atgggtcccc attctgacca ggtgtagcac agtggcactg tgggcgtgcc tgtgagctac 1260
acagactgtg ctctctcacc cgtcggttgt tcagttgtga catcagatag cttcaaaaaca 1320
aaagacagct tttagaactgc aaaaagtagg gaatacacaag gatgatgca 1380
cagttaatcc tagaaaataa ctctgggatt cccaaattga ctcttcgtag gcgtcatgat 1440
agcagcagca aaacaaatga ccaagagaat gatggaatga actcttccaa aataagcatc 1500
aagttaagca aagaccatga caacgataac aatctctatg tagcaaagct taataatgga 1560
tttaactcag gatcaggcag tagttctaca aaattaaaaa tccagctaaa acgagatgag 1620
gaaaataggg ggtctttatac agaggggctt catgaaaatg ggggtgtgctg cagtgatcct 1680
ctttctctct tggagtctcg aatggagggt gatgactata gtcagtatga ggaagaaagt 1740
acagatgatt cctcctcttc tgagggcgat gaagaggagg atgactatga tgatgacttt 1800
gaagacgatt ttattcctct tctccagct aagcgcttga ggttaatatg tggaaaagac 1860
tctatagata ttgacatttc ttcaaggaga agagaagatc agtctttaag gcttaatgcc 1920
taagctcttg gtcttaactt gacctgggat aactacttta aagaaataaa aaattccagt 1980
caattattcc tcaactgaaa gtttagtggt agcacttcta ttgtcccttc acttatcagc 2040
atactattgt agaaagtgt cagcactact actcaattct taagtctgat ttgtgcaaat 2100
ttttatcgta ctttttaaat agccttctta cgtgcaattc tgagttagag gtaaagcct 2160
gttgtaaaat aaaggctcaa gcaaaatttg acagtgtag caactttcca cacaggagct 2220
tgaaaacagt aatgtggcta cacagttttt ttaactgtaa gagcatcagc tggctcttta 2280
atatatgact aaacaataat ttaaaacaaa tcatagtagc agcatattaa gggtttctag 2340
tatgtctaata tcaccagcaa tgatcttttg ctttttgatt tatttgctag atgtttcccc 2400
cttggagttt tgtcagtttc acactgtttg ctggcccagg tgtactgttt gtggcctttg 2460
ttaatatcgc aaaccattgg ttgggagtca gattggtttc ttaaaaaaaa aaaaaaa 2517

```

<210> 40

<211> 1698

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1503131CB1

<400> 40

```

gactacgctt gctgggtttgc ggccgggtctt ggatgaagcg gcggccgctg tgagagcgtg 60
gggaagggtg ggggtgagggg gcgaggccgc agctaggggc gcgaaactct cctccccctg 120
gccccaccgc gtgggacggg gtgaactgtg tctcggaggg atgtcagcct tctctgaggc 180
ggcgctggag aagaagctgt cggagtgtag caactcgcag cagagcgtgc agaccttgtc 240
cctgtggctc attcaccacc gtaaacactc gcggcccatc gtcaccgtgt gggagcggga 300
gctgcggaaa gccaaaccaa acaggaagct tacttttctc tacctagcca atgatgtcat 360
acagaacagc aagaggaagg ggccagagtt tacaaaagat tttgcaccag ttatagtggg 420
ggcttttaag catgtttcaa gtgaaactga tgaaaagtgt aagaagcacc ttggaagagt 480
gttatcttat tgggaagaaa ggtctgttta tgaaaatgat gtattagaac aacttaaaaca 540
agctctgtat ggtgataaga agcctaggaa gcgaacttat gaacagataa aggtggatga 600
aaatgaaaac tgttctcttc tgggatctcc aagtgaacca ccacagactc tagatctcgt 660
tagagcatta caagatctgg aaaatgcagc ctcaggtgat gcagcagttc atcagaggat 720
agcttcttta cctgttgaag tccaagaagt atctctatta gatataataa cagataaaga 780
atctggagaa aggcctttcca aaatggtaga ggatgcgtgt atgttgctgg cagattacaa 840
tggcagattg gcggcagaaa tagatgatag aaagcaactc actcgaatgt tagcagattt 900

```

tcttcgttgt	caaaaggaag	cccttgcaga	gaaagagcat	aaattggaag	agtacaagcg	960
caagctagcc	agagtttccc	tggtgcgcaa	agaactcagg	tcccggatcc	agagcctgcc	1020
agacttatct	cgattgcccc	atgtcactgg	cagccacatg	cacctgccct	ttgcgggaga	1080
catctacagt	gaagattgat	ggaccagcct	ctttccaggt	cccaggactt	tgcaagagat	1140
ggagacaggt	taggtggata	gtcctgtagt	gttatttttg	tatatgtgtg	agaaagaaac	1200
actaacaana	ggaacaacga	gtaattttata	aaattgttta	aaatgttggt	ttgtttacta	1260
ttttattggg	aaattaacat	gacttaattt	gaacaaagtg	atgattttgt	tgtattaata	1320
agctcacaaa	tagtgattat	tttcaaaaga	tctttttgta	aatttttttt	tttgatccaa	1380
ggccttgtag	gaaatctctt	gtttgtattt	cttcagggtat	tttcaattgt	ttttctttat	1440
tttcttcagc	atttaatacac	tgtatactat	gtaattcctg	aaggggaaga	actaattagg	1500
agttgattga	gactttatta	tggtatagtt	aggacttgat	tgtagaaggg	actaaatgtt	1560
ccaacataat	ctttaccctt	ctcccaaac	aaaatatcaa	ttctgtgtct	ctacatgcct	1620
ttgattctct	agaggtcggt	aatttaacat	aaaacaggga	tgtcactttg	aggggcagat	1680
gaaagtctag	agacctgt					1698

<210> 41

<211> 2250

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1594803CB1

<400> 41

gctcgcagcg	catccccgcg	cacaggttcg	tgctggccgt	gggcagcgcg	tctttgatgc	60
catgttcaac	gggggaatgg	ccacaacatc	cacggagatt	gagctgcccg	acgtggaacc	120
cgtgccttc	ctcgcactgc	tcaagtttct	ctactcggac	gaggtgcaga	ttggcccggga	180
gacggtgatg	accacgctat	acaccgcaa	gaagtacgcg	gtgccagcgc	tcgaggccca	240
ttgcgtggag	ttcctgaaga	agaacctgcg	agccgacaac	gccttcatgc	tgctcacgca	300
ggcgcgactc	ttcgatgaac	cgcagctggc	cagcctgtgc	ctggagaaca	tcgacaaaaa	360
cactgcagac	gccatcaccg	cggagggcct	caccgacatt	gacctggaca	cgctgggtggc	420
tgtcctggag	cgcgacacac	tgggcatccg	tgaggtgcgg	ctgttcaatg	ccgttgtccg	480
ctgggtccgag	gccgagtgtc	agcggcagca	gctgcaggtg	acgccagaga	acaggcggaa	540
ggttctgggg	aaggccctgg	gcctcattcg	cttcccgtc	atgaccatcg	aggagttcgc	600
tgcaggtccc	gcacagtcgg	gcctcctggt	ggaccgcgag	gtggtcagcc	tcttctcgca	660
cttcaccgtc	aaccccaagc	cacgagtggg	gttctatgac	cggccccgct	gctgtgcgg	720
tggaagggag	tgcagcatca	accgcttcca	gcaggtggag	agtcgctggg	gctacagcgg	780
gaccagtgac	cgcacaggt	tctcagtcaa	caagcgcac	ttcgtgggtg	gatttgggct	840
gtatggatcc	atccacgggc	ccaccgacta	ccaagtgaac	atccagatta	ttcacaccga	900
tagcaaacacc	gtcttggggc	agaacgacac	gggcttcagc	tgcgacggct	cagccagcac	960
cttcgcgctc	atgttcaagg	agccggtgga	ggtgctgccc	aacgtcaact	acacggcctg	1020
tgccacgctc	aaggggccag	actcccacta	cggcaccaaa	ggcctgcgca	agggtgacaa	1080
cgagtcgccc	accacggggc	ccaagacctg	cttcaccttt	tgctacgcgg	ccgggaacaa	1140
caatggcaca	tccgtggagg	acggccagat	ccccgaggtc	atcttctaca	cctaggctgc	1200
ccgacaccca	caccgccttc	cctccgtggg	gatagccgca	gccccaggcc	atcatctgct	1260
gctggggccc	ccccaccacg	cgggtgccag	cccagtgtcc	cccaggccgt	ctgtccactc	1320
catgccacct	ttctcagcat	caggacgggg	ttgccctgtg	ttcaccacga	gtgtggctgc	1380
tggatcaggg	cagccggggg	ggtggccagg	ccagtggcca	ggccctgtgg	agacaatccc	1440
tcaggactag	ggacagggct	gtgccggcct	gggccagggc	ccacggaccc	gcagctcagg	1500
gcgcctgccc	acgtcgtctg	cggcggtgct	gcccggggcg	tccctcgctg	ctcttactg	1560
cacattgcaa	tgcatttgcg	attcccattt	ctctgttagg	agccagcctg	ggtggcgctg	1620
ctcccagagc	cgtgggtccc	agaccttgcg	ttccttttgt	tcctgtccgt	ttatcaggac	1680
acggggcccca	cctgtcacgt	gcccaggggc	acccaagccc	agcctgcggg	gcgttcccac	1740
tgcttggatg	ccggcttgag	ttctgcgcac	gcaggattca	gtgtggggac	ggccccctgc	1800
ggataggcct	agccctggcc	caggtggtga	gcggtttgca	gtgtccgttc	tcatecacct	1860
gatgggcccc	gataaaggcc	cccgtgtgct	agcctccctg	gacggccctc	gcggtccctg	1920
cagcccaaga	tgggactcag	accctgtgcc	ccagagctcc	cctgcccagc	aatggggccc	1980
cagccggccc	cgaccgggtc	caggagcact	gctgcgctgt	acatactgtt	gccctagccc	2040
acctggtgcc	gtgggagcca	cccccagggt	ctgggggac	agccccctcc	cactccggcc	2100
acgccccac	ccaccccgcg	tgtttctgcc	ctgtgactcc	tggaacctgc	gtcctcccca	2160
aagccatggg	aggggtgtcc	tcctcagacc	atgccccag	atgatttttt	taaataaaga	2220
aacaaatgca	cctgcaaaaa	aaaaaaaaaa				2250

<210> 42

<211> 1082

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1736129CB1

<400> 42

ggcaaatccg	ggatctcggc	tccgagaggc	tcattctggca	aaagggcgcg	gaaaccacgg	60
gggccctgag	actgagtggg	ccctatccca	cttctctctg	gggcggcgct	gtagccagcg	120
gctgacaggc	gcacgaatag	gcagccctct	gctgtaagga	ggaaaactga	ggcctgggag	180
caggaacctg	taggcagcgc	ttgagggtag	cgggatatga	gctgcaagcg	cgcgtgggag	240
gcgggggctc	tgggcgggaac	aaaaatcaca	ggatgtcaga	ggatgtttcc	cgggaagaac	300
tgggataaag	gaagggtccc	agcaccatgg	aggacccgaa	ccctgaagag	aacatgaagc	360
agcaggattc	acccaaggag	agaagtcccc	agagcccagg	aggcaacatc	tgccacctgg	420
gggccccgaa	gtgcacccgc	tgccctcatca	ccttcgcaga	ttccaagtcc	caggagcgctc	480
acatgaagcg	ggagcaccca	gcggacttcg	tggcccagaa	gctgcagggg	gtcctcttca	540
tctgcttcac	ctgcgcccgc	tccttccccct	cctccaaagc	cctaataacc	caccagcgca	600
gccacgggtc	agccgcgaag	cccaccctgc	cgggttgcaac	cactactgcc	cagccccact	660
tcccttgctc	tgactgtggc	aagacccttg	ggcaggctgt	ttctctgagg	cggcacccgc	720
agatgcatga	ggtccgtgcc	cctcctggca	ccttcgcctg	cacagagtgc	ggtcaggact	780
ttgctcagga	agcagggtcg	catcaacact	acattcggca	tgcccggggg	gagctctgag	840
tgcagcttaa	gcctctccac	ggtgacgggt	ggctctgtgg	ctggtaggac	tcacccatga	900
tatgggggtg	aggaaactcg	ggggccctga	aggatttgct	tccttccccct	gggaaggcag	960
agggtcttta	ataaagagga	cccagaagat	tccttatttag	agcttcagtc	tttggagcac	1020
acagggcctt	cgtgagacag	tgaaatcaga	taataatgag	atctttttgtt	aaaaaaaaaa	1080
aa						1082

<210> 43

<211> 570

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1874312CB1

<400> 43

gttcgcgtgt	gctgttttttc	cgctcatggct	cgcactaagc	aaactgctcg	gaagtctact	60
ggtggcaagg	cgccacgcaa	acagttggcc	actaaggcag	cccgcaaaag	cgctccggcc	120
accggcgggc	tgaaaaagcc	ccaccgctac	cggccgggca	ccgtggctct	gcgcgagatc	180
cgccgttatc	agaagtccac	tgaactgctt	attcgtaaac	tacctttcca	gcgcctgggtg	240
cgcgagattg	cgcaggactt	taaaacagac	ctgcggttcc	agagctccgc	tgtgatggct	300
ctgcaggagg	cgtgcgaggc	ctacttggtg	gggctatttg	aggacactaa	cctgtgcggc	360
atccaacgcc	aagcgcgtca	ctatcatgcc	caaggacatc	ccactcacc	gccagcatcc	420
gcggaagaga	gggcgggtgat	tactgtgggt	ctctcttgcc	ggtccaagca	aagggtcttt	480
ttcagggcc	ccaccttttc	caaataaagt	tgctgtaaga	aaacccattt	atggcccaaa	540
agggattcct	ttggggggca	tttttctttt				570

<210> 44

<211> 1007

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1969301CB1

<400> 44

cccacgcgtc	cgcccacgcg	tccggtttgg	gtttcttcgc	ggctgctcaa	gatgaaccga	60
ctcttcggga	aagcgaaacc	caaggctccg	ccgcccagcc	tgactgactg	cattggcacg	120
gtggacagta	gagcagaatc	cattgacaag	aagatttctc	gattggatgc	tgagctagtg	180
aagtataagg	atcagatcaa	gaagatgaga	gagggtcctg	caaagaatat	ggtcaagcag	240
aaagccttgc	gagtttttaa	gcaaaagagg	atgtatgagc	agcagcggga	caatcttgcc	300
caacagtcac	tcaacatgga	acaagccaat	tataccatcc	agtctttgaa	ggacaccaag	360
accacggttg	atgctatgaa	actgggagta	aaggaaatga	agaaggcata	caagcaagtg	420
aagatcgacc	agattgagga	tttacaagac	cagctagagg	atatgatgga	agatgcaaat	480
gaaatccaag	aagcactgag	tcgcagttat	ggcaccocag	aactggatga	agatgattta	540

```

gaagcagagt tggatgcact aggtgatgag cttctggctg atgaagacag ttcttatattg 600
gatgaggcag catctgcacc tgcaattcca gaaggtgttc ccactgatac aaaaaacaag 660
gatggagttc tgggtggatga atttggattg ccacagatcc ctgcttcata gatttgcac 720
attcaagcat atcttgtaaa acaaacacat attatgggac taggaaatat ttatctttcc 780
aaatttgcca taacagattt aggtttcttt cctttctttg aaggaaagt ttattacatt 840
gctcttttat tttttccatt aagagactca ttgcttgga aatgctttct tcgtactaaa 900
atgtgattcc tttttttctt atgaaaaacg aactcagttt aaaagtattt ttagctcgta 960
tgacttggtt tcattcatta ataataattt gagataacca agggaat 1007

```

```

<210> 45
<211> 1622
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1986873CB1

```

```

<400> 45
ttaaccaa at gatgcaaa ac agcacctttt ttcatagcac gacagcaaca aattttgcaa 60
taattactgt ccttgtccac gcttcatgta ttttgcaatc tcctcttcgg agataaaagc 120
ttctttttct tcagtaaaat ggacatgact agtccagagc agtctagaaa tgtgctacag 180
tttactgaag aaaaagaagc ttttatctct gaagaggaga ttgcaaaaata catgaagcgt 240
ggaaaaggaa agtattattg caaaatttgt tgctgtcgtg ctatgaaaaa aggtgctgtt 300
ttgcatcatt tgggttaataa gcataatgtt catagccctt acaaatgcac aatctgtgga 360
aaggcttttc ttttggaaatc tctccttaaa aatcatgtag cagcccatgg gcaaagttaa 420
cttaaatgtc cacgttgtaa ttttgaatca aatttcccaa gaggttttaa gaaacattta 480
actcattgtc aaagccggca taatgaagag gcaaataaaa agctaagga agctcttgaa 540
ccgccactgg aggagcagca aatttgataa cacagtgtga atatttgctc tacaaagggtg 600
tttgttgga ccattctttg taagtatagc ttatcagata gcatagttgg atcagtagat 660
gacatgtatg gtgtaccgtg tttcactgtc tcagtgtgtg tactaagaat gagcatttga 720
tcattttttt ctggtctctg tctatgtgac tatcttgtaa gtcaataaat ttctgtatag 780
tccagatgga ttaaacttct catttctttt aaatatgtat gaataataat acaagggaagt 840
aggcattcca ttttaataatc aagagcaagt tgtactcaaa gcattcagtt aaagtgtatc 900
tgtgtgtgga actaattttca gacaatagaa aatattagtt gaaatgttta agaattaggc 960
atgaaaaata aatttgagaa attttgtttc cttacatgta tttttaaatc ataagagtta 1020
ttttctatct gatgtaaaat tagtttataa atcttaatca gcttctagat gtttatttagc 1080
ttttatgtca tgaaatgttg gagtctcagg gttgctgatt ttctgcta at gggaaaaatt 1140
gactaagtct ttaaaatagt ttgcagcctt ctcccacagg agacaagtga aagataagtg 1200
tgatttttaga tctttctttg ccatagttgt tttcagtggg gtcttccatt ctgtatctta 1260
ccctaagatc tgggtcttcc ctccccatgc ccaccgccct cccaccgccct cccagctcac 1320
actaatagat gattcttaat tgccaaatgt gttagagttt gtatatccta ctctggggcc 1380
ttacatgtcg cctgttgggg cttaagacca ggttgataag taggaactga aagtcttcca 1440
gattcacagt agaaaatttt atagacattt ctgtttaaaga aatataatga ttttatgttt 1500
ttcaattatg ttactgtaaa taccttgtac ctgttcatgg attattttat tctaaaaaat 1560
tttgtcaaat gtgtatcaac caaattaaaa agaaagggtt tcatgtcaaa aaaaaaaaaa 1620
aa 1622

```

```

<210> 46
<211> 2047
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2010820CB1

```

```

<400> 46
tggtgaaaga ctactccaca tcattataga tgataataag agaacacaaa ttgttacaga 60
aattatctca gagattcggg cgccatttgt tactgttggg gtttaataacg atccagctga 120
tgtaagatag aaagaactca agatggctga aataaaaagt aagcttatcg aagccaaaga 180
agcttttgga aattgcatta cttacagga ttttaatcgg gcatcagaat taaaagaaga 240
aataaaaagca ttagaagatg ccagaataaa ccttttgaaa gagacagagc aacttgaaat 300
taaagaagtc cacatagaga agaatgatgc tgaaacattg cagaaatgtc ttattttatg 360
ctatgaactg ttgaagcaga tgtccatttc aacaggctta agtgcaacca tgaatggaat 420
catcgaatct ttgattcttc ctggaataat aagtattcat cctgttgtaa gaaacctggc 480
tgttttatgc ttgggatgct gtggactaca gaatcaggat tttgcaagga aacacttcgt 540

```

```

attactattg caggttttgc aaattgatga tgtcacaata aaaataagtg ctttaaaggc 600
aatctttgac caactgatga cgttcgggat tgaaccattt aaaactaaaa aaatcaaaac 660
acttcattgt gaaggtacag aaataaacag tgatgatgag caagaatcaa aagaagttga 720
agagactgct acagctaaga atgttctgaa actcctttct gatttcttag atagttaggt 780
atctgaactt aggactggag ctgcagaagg actagccaag ctgatgttct ctgggctttt 840
ggtcagcagc aggattcttt ctctgtcttat tttgttatgg tacaatcctg tgactgaaga 900
ggatgttcaa cttcgacatt gcctaggcgt gttcttcccc gtgtttgctt atgcaagcag 960
gactaatcag gaatgctttg aagaagcctt tcttccaacc ctgcaaacac tggccaatgc 1020
ccctgcatct tctccttttag ctgaaattga tatcacaat gttgctgagt tactttaga 1080
tttgacaaga ccaagtggat taaatcctca ggccaagact tccaagatt atcaggcctt 1140
aacagtacat gacaatttgg ctatgaaaat ttgcaatgag atcttaacaa gtccgtgctc 1200
gccagaaatt cgagtctata caaaagcctt gaggttcttta gaactcagta gccatcttgc 1260
aaaagatcct ctggttctat tgaatgagat tctggagcaa gtaaaagata ggacatgtct 1320
gagagctttg gagaaaatca agattcagtt agaaaaagga aataaagaat ttggtagcca 1380
agctgaagca gcacaggatg ccaccttgac tacaactact ttccaaaatg aagatgaaaa 1440
gaataaagaa gtatatatga ctccactcag ggggtgtaaaa gcaacccaag catcaaagtc 1500
tactcagcta aagactaaca gaggacagag aaaagtgaca gtttcagcta ggacgaacag 1560
gaggtgtcag actgctgaag ccgactctga aagtgatcat gaagttccag aaccagaatc 1620
agaaatgaag atgagactac caagacgagc caaaaccgca gactagaaa aaagtaaaact 1680
taaccttgcc caatttctca atgaagatct aagttaggaa agacgatgga ggtggaaatcc 1740
tttaagatta tgtccagtta tttgctttaa taaagaagaa gttacccttg tcaaaatcag 1800
aacaaacctg atgtctttct gaagattttt tgctgtgcgc ttccacgtta ctttggcctg 1860
tattaaagca gttagcagc atcagttatt atagtccaga aaaagtgtgc atcagtcag 1920
cacacagatt tatcccaatc tgaaggtggg ctaggaatct catttttaaa tagtctctcc 1980
aagtgattcc tatggactct ttaagtttaa atcatgtcct tatggaaaac ttacagtgtt 2040
actagct
2047

```

<210> 47

<211> 1817

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2013818CB1

<400> 47

```

caacaggaag catcagcaga tgttgctact cctaagatgc cagggcagtc agtcaggaag 60
aaaactagga aggcaaaaaga aatttctgaa gcttctgaaa acatctattc tgatgtcaga 120
ggactatctc agaaccagca aatacctcaa aattctgtta cgcctaggag aggaaggaga 180
aagaaagaag ttaactcagga catactagaa aacaccagtt ctgtggaaca agaattacag 240
atcactacag ctagggaatc aaaaagatta aaatcatctc agctgttggg accagcagtt 300
gaagaaacta ctaaaaaaga agttaagggt tcactctgtta caaaaaggac tcctagaaga 360
attaaaagat ctgtagaaaa tcaggaaaagt gttgaaatta taaatgatct aaaagttagt 420
acggttaacaa gtcctagcag aatgatcaga aaattgagaa gtactaattt agatgcttct 480
gaaaatacag gaaataagca agatgataaa tccagtgaca agcagctgcg tattaacatc 540
gttagaaggg tcagagggag agaagttagt ccacagatg tgagagaaga ctccaacctt 600
gagtcatctc agttgactgt tcaagcagaa tttgatattg ctgccatacc tagaaaacgt 660
ggtagaccaa gaaaaatcaa tccatctgaa gatgtaggat ctaaggctgt taagggaagag 720
agaagcccca agaagaaaga agctcccagc attagaagga gatctacaag aaatacccca 780
gctaaaaagt gtaaatgttg tgttggaaaa ccagctttag gaaaatccat tttagtgcca 840
aacgaggaac tttcgatggg gatgagctct aagaaaaaac ttacaaaaaa gactgaaagt 900
caaagccaaa aacgttcatt gcaactcagta tcagaagaac gcacagatga atgacacat 960
aaagaaacaa atgagcagga agaaagattg ctgcgccacag cttccttcac taaatcatcc 1020
cgcagcagca ggactcgggc tagcaaggcc atcttgttgc cggaccttct tgaaccaaac 1080
aatgagcctt tattttctcc agcgtcagaa gttccaagga aagcaaaagc taaaaaata 1140
gaggttcctg cacagctgaa agaattagtt tcggatttat cttctcagtt tgtcatctca 1200
cctcctgctt taaggagcag acaaaaaaac acatccaata agaacaagct tgaagatgaa 1260
ctgaaaagat atgcacaatc agtagaaaact ctgggaaagc caaaagcgaa acgaatcagg 1320
acgtcaaaaa caaaaaagc aagcaaaaac ctagagcccag ctgacggagt caacagcaa 1440
cccatagaaa ttcggctgat gacaggaaca ggtcttggaa ggaacagaaa gaaactgtct 1500
ccctatccaa agcaaatatt acgcagaaaa atgctgtaat ttcttgggaa gattttaatg 1560
tacaccttat tgaagtca tcagaatagt gtggattatt aaatatctag tttggaagaa 1620
aataatttat ataaattatt gtaaattttt atgtaaacag aaggtcttca ataagtaaa 1680
taactccata tggagtgatt gtttcagtc aggcaatttt tctattttat attaagactt 1740
catacattta tatatgtaaa tatggcttat taatggaatg ttaataaaaa tgtatacttc 1800

```

acagtcaaaa aaaaaaa

1817

<210> 48

<211> 700

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2302032CB1

<400> 48

cccgcctcgc	agagttggga	gaaggcaggg	tggggggtgt	ggaaaaataa	aaggaaaagt	60
ccttgaccca	tgtagatcag	cgtcccccac	tttggcatcc	cggccggccg	gggacctccc	120
agtctgcggc	catgaacgcg	agcagcagag	gcgagagctt	cgcgggctcg	gtgcaaattc	180
caggtggcac	aacggtgctg	gtggagctga	ctcccgacat	ccatatctgc	ggcatctgca	240
agcagcagtt	taacaacctg	gatgcctttg	tagctcacia	gcaaagtggc	tgccagctga	300
caggcacatc	cgcagcagcc	cccagcagcg	tccagtttgt	atcggaggaa	acagtgcctg	360
ccacccagac	tcagaccacc	accagaacca	tcacctcgga	gacccagaca	atcacaggta	420
cagctggagc	atgggggagt	cggccagaat	tggcctggct	gtgtctcaaa	cacgtccatg	480
gaacttggtt	aaaatataga	ttcccagagc	ccacccccta	gggttgctga	ttcattatct	540
ttaacaagct	ccaccaggaa	gccagctgcc	cagccaggtt	tgggagctgc	taggttaaaa	600
cattaactgt	tagcagttat	aatgtttgtc	acgttacaaa	aaaaacagac	aaaaaagaga	660
tttctgaaac	aaaaatctat	atactatagc	aaaaaaaaaa			700

<210> 49

<211> 2704

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2326109CB1

<400> 49

cggggcgagg	gcggaggcag	aggtggaggg	gctttgaaa	ccgtcattgt	gtggcactgg	60
acgggggaat	aagagacccg	gtttctaat	cctgctctgt	cactaacaag	ctgcgtgac	120
ttcggaaaaa	agtaatcctc	ttggaaagct	ggaaacatag	tccgaagctt	aagtgaata	180
gatacatgtt	agaagacatg	aaacaaggag	aaattctaat	agtcacgtgc	ctgagcagaa	240
atctcgagtt	gattggaggc	gaactaaaag	aagtagtatc	tcacaattac	ttgatagtga	300
tgaagagctt	gatagtgaag	aatttgatag	tgatgaagag	cttgatagt	atgaaagt	360
tgaaaaatgat	gaagagcttg	atagtaacaa	gggacctgat	tgtaataaaa	caccaggaag	420
tgaaagagag	ctcaacttaa	gtaaaattca	aagtgaagga	aatgacagta	agtgtctcat	480
taactctggc	aacggttcaa	catatgaaga	agaaacgaac	aaaatcaaac	ataggaatat	540
tgacttacaa	gatcaggaaa	aacattttaag	tcaagaggat	aatgatctca	acaaacaaac	600
tggaacaaata	atagaggatg	atcaggaaaa	acatttaagt	caagaggata	atgatctcaa	660
caaacaaaact	ggacaaaata	tagaggatga	tttagaagaa	gaagacatca	agcgaggaaa	720
aagaaaaagg	ctatcctctg	tgatgtgtga	cagtgtgag	agtgtgaca	gcgatatcct	780
agttagaaaa	gtaggtgtta	aacgtccccg	tagagtgggt	gaagatgaag	gttcttcagt	840
ggaaatggag	caaaagactc	ctgaaaaaac	attagctgca	caaaagcgag	aaaaacttca	900
gaagctcaaa	gaactctcaa	aacaaagatc	tcgtcagaga	cgcagtagtg	gtagagattt	960
tgaggactct	gaaaagggaat	cttgcccaag	cagtgtgaa	gttgatgagg	aggaagaaga	1020
ggataattat	gaatctgatg	aagatggaga	tgattatatt	atcgatgact	ttgtagtgc	1080
agatgaggag	ggtgatgaag	agaataaaaa	ccaacaagga	gaaaaattga	ctacatcaca	1140
actgaaatta	gtaaaacgga	attctcttta	ttcttttagt	gaccactata	ctcattttga	1200
aagagttgtg	aaggctcttc	tgatcaacgc	tttagatgaa	tcttttctgg	gaacattata	1260
tgatggcaca	aggcaaaaat	catatgcaaa	agatatgcta	acatctcttc	attatttgga	1320
taaccgcttt	gttcagcctc	gtctagagag	cttggtatct	agaagtcgtt	ggaaagagca	1380
atataaggag	cgagtagaaa	attattctaa	tgtaagtatt	catttgaaag	atcctgaaaa	1440
ctgttcctgc	caggcttggtg	gactgcatcg	ctactgtaaa	tattcagtgc	atattatcagg	1500
agagttgtat	aacaccagga	ccatgcaaat	agataatttc	atgtcacatg	ataaacaggt	1560
gttcaactgtt	ggcagaattt	gtgccagccg	taccagaatt	tatcataaac	tgaaacattt	1620
taaattcaaa	ctataccagg	aatgttgcac	cattgcaatg	acagaagaag	ttgaagatga	1680
acaagttaaa	gaacacagtg	aaagaatttt	caggcggtca	aaagaaaatg	gctggattaa	1740
ggagaaatat	ggtcaacttg	aagaatatct	caattttgcg	gattactttc	aagaagagaa	1800
gtttgagttg	taacacattt	ccatcagaga	agatttttta	aattcctgta	aatgtgaaga	1860
tcatgattct	tgttttttctg	tatcatgtga	catgtttgta	catttttatct	atatcttcat	1920


```

ggcaaaatat tttgtttaaa acatgattat cttaattcct gtgaagtggc actctacact 1980
ctaattaaac ttttatctga tgtacataaa agcattatct taatttttaa gtctgtaa 2040
atattttaaa gttatataat aatggcttta taacagatga ctgtcaagtg aatgagctgt 2100
tgatatcctg tcagtttagt caaaatatat tgtatcttaa aaatgtattt agactaacgt 2160
ctacatgtat ttttacggaa ttctgtcca gatctgttta ttcttttaca gtattaaatg 2220
attttcaacta atatatTTTT tactgctatc atctaaatca gtgggccatg gaagaaacct 2280
catatatcat tcagttcaat cccattgctt cagaatcaga tcatagcata tctattctaa 2340
gtagaggaaat gcTTTTTtcc ttttcttgaa agtttctctt tcaaatagat ttcaggctc 2400
tagagtcttt caacctcac attcaggaat aattttatgt aaattttcat gctttataat 2460
gttcttactt ttttctatct aattttgtct atatatgtat gattaagatg tattttat 2520
atttttatcc atagctTTTT ccatataagt atgtatctta gggtcagaac tgctgaaaga 2580
gacaaactca gccaaaaaca cttggaaagc atattttgggt atctgcattg ctttgcata 2640
ctaaattttc ccataaagta ataaagtaaa atggttgag agtgaacaaa aaaaaaaaaa 2700
aagg

```

<210> 50

<211> 3086

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2354751CB1

<400> 50

```

cgccggcgtc cgcccagccg acccctctgg gccggggcgc acggctcccc gaactgggca 60
gccgggggac agaagtccgt cctaggcccc ccaggctctg accttctttc ccaggatgag 120
gtggggccac catttgccca gggcctcttg gggctctggg tttagaagag cactccagcg 180
accagatgat cgtatccctt tcctgatcca ctggagttgg ccccttcaag gggagcgctc 240
ctttggggcc cctagggcct ttatacgcca ccacggaagc tcggtagata gcgctcccc 300
acccgggagg catggacggc tgttccccag cgctcttgca actgaagcta tacagcggca 360
ccgcgggaac ctggctgagt ggttcagccg gctgccagg gaggagcgcc agtttgccc 420
aacctttgcc ctagacacgg tccacgttga cctgtgatc cgcgagagta cccctgatga 480
gctacttcgc ccaccgcgg agctggccct ggagcatcag ccaccacagg ccgggctccc 540
ccactggcc ttgtctcagc tctttaaccc ggatgcctgt gggcgccggg tgcagacagt 600
ggtgctgtat gggacagtgg gcacaggcaa gagcacgtg gtgcgcaaga tggttctgga 660
ctggtgttat gggcggtgc cggccttcga gctgctcat ccttctctgt gtgaggacct 720
gtcatccctg ggccctgccc cagcctccct gtgccaactt gtggcccagc gctacacgcc 780
cctgaaggag gttctgcccc tgatggctgc tgctgggtcc cacctcctct ttgtgctcca 840
tggcttagag catctcaacc tcgacttcgg gctggcaggc acgggacttt gtagtgacct 900
ggaggaaccg caggaaccag ctgctatcat cgtcaacctg ctgcgcaaat acatgctgcc 960
tcaggccagc attctggtga ccactggccc ctctgccatt ggccgtatcc ccagcaagta 1020
cgtggggcgc tatggtgaga tctgcggttt ctctgatacc aacctgcaga agctctactt 1080
ccagctccgc ctcaaccagc cgtactgcgg gtatgccgtt ggcggttcag gtgtctctgc 1140
cacaccagct cagcgtgaec acctggtgca gatgctctcc cggaacctgg aggggcacca 1200
ccagatagcc gctgcctgct tctgcccgtc ctattgctgg ctggtttgtg ccaccttgca 1260
cttctctcat gccccacgc ctgctgggca gaccttaca agcatctata ccagcttccc 1320
gcgcctcaac ttcagcgggg aaacctgga cagcactgac ccttccaatt tgtccctgat 1380
ggcctatgca gcccgaaaca tgggcaagtt ggcctatgag ggggtgtcct cccgcaagac 1440
ctacttctct gaagaggatg tctgtggctg cctggaggct ggcatacagg cggaggagga 1500
gtttcagctg ctgcacatct tccgtcggga tgcctgagg ttttctctgg ccccatgtgt 1560
ggagccaggg cgtgcaggca ccttcgtgtt caccgtgccc gccatgcagg aatacctggc 1620
tgccctctac attgtgctgg gtttgcgcaa gacgacctg caaaagggtg gcaagggaag 1680
ggctgagctc gtgggcccgt ttggggagga cgtcagcctg gtactgggca tcatggccaa 1740
gctgtctgct ctgcccgtct tgccctgctg cttcaacctg atcaagggtg ttccacgagt 1800
gtttgggcgc atggtgggta aaagccggga gccggtggct caggccatgg tgctggagat 1860
gtttcgagag gaggactact acaacgatga tgttctggac cagatgggag ccagtatcct 1920
gggcgtggag ggcccccgcc gccaccacaga tgagccccct gaggatgaag tcttcgagct 1980
cttccccatg ttcatggggg ggcttctctc tgcccacaac cgagctgtgc tagctcagct 2040
tggctgcccc atcaagaacc tggatgcccc ggagaatgcc caggccatca agaagaagct 2100
gggcaagctg ggccggcagg tctgcccccc atcagagctc cttgaccacc tcttcttcca 2160
ctatgagttc cagaaccagc gcttctccgc tgaggtgctc agctccctgc gtcagctcaa 2220
cctggcaggt gtgcgcatga caccagtcaa gtgcacagtg gtggcagctg tgctgggcag 2280
cggaaggcat gccctggatg aggtgaactt ggcctcctgc cagctagatc ctgctgggct 2340
gcgcacactc ctgcctgtct tctgcgtgac ccggaagctg ggcttgcaac tcaacagcct 2400
gggcccctgag gcctgcaagg acctccgaga cctgttgctg catgaccagt gccaaattac 2460
cacactgcgg ctgtccaaca acccgctgac ggaggcaggt gttgccgtgc taatggaggg 2520

```

gctggcagga	aacacctcag	tgacgcacct	gtccctgctg	cacacgggcc	ttggggagca	2580
aggcctggag	ctgctgggctg	cccagctgga	ccgcaaccgg	cagctgcagg	agctgaacgt	2640
ggcgtacaac	ggtgctggtg	acacagcggc	cctggccctg	gccagagctg	cccgggagca	2700
cccttccctg	gaactgctac	agtgaagtcc	gtccctgggc	ccattgcccc	cagcccttca	2760
gactacttcc	ctctctcaac	tgtgctcctc	caatcctagg	gagtgttctt	gggcctgggc	2820
ctgggtgcctg	tcctttatct	ctgggtttctg	actgagcctt	gtcaacctag	ctctgagcca	2880
gcccaccccc	tctgtgactg	gcgcttaaac	tgcaccaatt	ccctaccacc	atttctccta	2940
aagccctact	gtctctggct	ccactgaatt	ctagtatcag	ggtcggctgt	ggtgggtcac	3000
gtctgtaatc	ccagcacttc	ggggggccaa	ggcgggcaca	tcacctgagg	tcaggagttc	3060
aagaccagtc	tggccaacat	ggtgaa				3086

<210> 51

<211> 676

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2378058CB1

<400> 51

ttcgtttttg	tgcgcgcgagg	ttttggtttg	tgaggatcgg	cgagtggcgc	ccaccatctc	60
tgcttgctga	aaagctgcag	aggccgccag	gagccacgc	acctgagact	tttaccttta	120
cccagaaagg	aataaaagag	gtcagatgat	gacagctgtg	tccttaacaa	ccaggcccca	180
ggaatcagtg	gcttttgagg	acgtggctgt	gtacttcaat	acgaaggaat	gggccatcat	240
ggtgcctgcc	gagagggcct	tgtacaggga	tgtgatgctg	gagaactatg	aggctgtggc	300
ctttgtagt	cccccactt	ccaaaccagc	tttggctctt	catctggagc	aagggaaga	360
gtcctgtttc	acccagccac	agggagtcc	aagcaggaat	gactggagag	caggctggat	420
aggatacttg	gaactaagaa	gatatactta	cttagctaaa	gcagtgttac	gtagaatagt	480
atcaaaaatt	tttcgaaatc	gtcaatgctg	ggaagacagg	agaaaagctt	aattcttgac	540
atttaaatac	cagttttcca	agtaaggagt	tgatgtaaga	gccaccttaa	acgatgtcaa	600
atacacattt	tctttttctg	tagactggta	gatttaattg	ttttcattca	ttaaaataac	660
tcattttgat	gtacaa					676

<210> 52

<211> 2135

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> 1913

<223> a or g or c or t, unknown, or other

<220>

<221> misc_feature

<223> Incyte ID No: 2595747CB1

<400> 52

agtcccagcc	ttggaagcac	actgtgtaga	atttctcacc	aaacatctta	gggcagataa	60
tgcttttatg	ttacttactc	aggctcgatt	atttgatgaa	cctcagcttg	ctagtctttg	120
tctagatata	atagacaaaa	gcacaatgga	tgcaataagt	gcagaagggt	ttactgatat	180
tgatatagat	acactctgtg	cagttttaga	gagagacaca	ctcagtattc	gagaaagtcg	240
acttttttga	gctgttgtac	gctgggcaga	agcagaatgt	cagagacaac	aattacctgt	300
gactttttgg	aataaacaaa	aagttctagg	aaaagcactt	tccttaatcc	ggttccact	360
gatgacaatt	gaggaatttg	cagcaggtcc	tgctcaatct	ggaattttgt	cagatcgtga	420
agtggtaaac	ctctttcttc	attttactgt	caaccctaaa	ccccgagttg	aatacatgta	480
ccgaccaaga	tgctgtctca	ggggaaagga	atgctgcac	aatagattcc	agcaagtaga	540
aagccgctgg	ggttacagt	ggacgagtga	tcgaatcaga	ttcacagtta	atagaaggat	600
ctctatagtt	ggatttggct	tgtatggatc	tattcatggc	cctacagatt	atcaagtga	660
tatacagatc	attgaatatg	agaaaaagca	aaccttggga	cagaatgata	ccggcttttag	720
ttgtgatggg	acagctaaca	cattcagggt	catgttcaag	gaacccatag	agatcctgcc	780
caatgtgtgc	tacacagcat	gtgcaacact	caaagggtcca	gattcccact	atggcacaaa	840
aggattgaag	aaagtagtgc	atgagacacc	tgctgcaagc	aagactgttt	ttttcttttt	900
tagttccctt	ggcaataata	atggcacttc	aatagaagat	ggacaaattc	cagaaatcat	960
attttatata	taatttagca	ttataatata	tcttggttaa	ataataccat	acaatctagt	1020
gtcaaaaaaca	taaatggcca	caaaaaagta	gtttgagtgt	tatgaatatt	taaaattgta	1080

```

agataagaaa cagtttctta gagcagatag aaaaatgctt atttaaactt ttgcatgatt 1140
taaaaacaga ttttccattt tcttacaact ttaagagaaa agaactgggt ttaatgggtt 1200
aaaaaaaaagc acagcttttt caccttcac tttgtataatt tcatagattg gctgacttag 1260
ggtctttcaa tagtttggga attgaaagat tcttggtata tatagctagt ttgggtttgt 1320
ttttgtttta actattttga aggttaggtg agatgggcaa ataggcttaa ctattttgaa 1380
ggttggatga aaagagatgg gtcagtattc ctacagaatt cttattaact caaataacta 1440
aatttcagaa aattaagaag ctgactttat atttgggtgg ttgaagtatc ttgttggttag 1500
catttgtaat aatgctaaaa aaggcctaataaaaatgcccc agaaaatatt cagtgcattt 1560
atagagaagg atattttgta gtagtatagt aatgtgttat gtagtacagt tttaaagcta 1620
taaattggaat tttgtgtaaa ttcacaaaaa tgtgatataa acaggatcta agactggatt 1680
ccctgtcact aaactgcaca actatacctg tctctctgtg tgggggacac tgctgatgat 1740
tccaagatt taggatgatg aacggtgatg aacgcttggg tgaacagcca tcaactttta 1800
acattgggga ttaatccttt caacaggcaa ggagaacaag gatttaaaat tgactaacac 1860
aatttcttaa acaaacattg cttcatgacc aaatttgata agggaggatt canaacaggg 1920
tagttaaat cggggttg aaagataaa cgagttttca tgaccaacca ttttctgagg 1980
gggggggaac acggttaaca caatttcgac cattttggag ggggccggtt tacaatttaa 2040
atggccggtg ggtttaaaac gcggtgactg gggaaccatg ggggtaccaa cattggattt 2100
tttgaggaaa tccccttttt gcaagggggg tgatt 2135

```

<210> 53

<211> 1385

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2634391CB1

<400> 53

```

gccaacatt ccaagtcagg ggctcccaac aaatgataga ccaggcttcc ctgtaccagt 60
attctccaca gaaccagcat gtagagcagc agccacacta caccacaaaa ccaactctgg 120
aatacagtcc ttttcccata cctccccagt ccccgcgtta tgaacaaaac ctctttgatg 180
gtccagaatc acagttttgc ccaaaccaaa ccttagtttc cttctctggg gatcaaaggg 240
aatctgagaa tattgctaata cccatgcaga ctctctccag tgttcagcag caaaatgatg 300
ctcacttgca cagcttcagc atgatgccca gcagcgctg tgaggccatg gtggggcacg 360
agatggcctc tgactcttca aacacttcac tgccattctc aaacatggga aatccaatga 420
acaccacaca gttagggaaa tcactttttc agtggcagggt ggagcaggaa gaaagcaaata 480
tggcaaatat ttocaaagac cagtttcttt caaaggatgc agatggtgac acgttcttc 540
atattgctgt tgcccaaggg agaagggcac tttctatgt tcttgcaaga aagatgaatg 600
cacttcacat gctggatatt aaagagcaca atggacagag tgcccttcag gtggcagtg 660
ctgccaatca gcatctcatt gtgcaggatc tggatgaacat cggggcacag gtgaacacca 720
cagactgctg gggaagaaca cctctgcata tgtgtgctga gaagggccac tccaggtgc 780
ttcaggcgat tcagaaggga gcagtgggaa gtaatcagtt tgtggatctt gaggcaacta 840
actatgatgg cctgactccc cttcactgtg cagtcatagc ccacaatgct gtggtccatg 900
aactccagag aaatcaacag cctcattcac ctgaagtcca ggagctttta ctgaagaata 960
agagtctggt tgataccatt aagtgcctaa ttcaaatggg agcagcggtg gaagcgaagg 1020
cttacaatgg caacactgcc ctccatggtg ctccagctt gcagtatcgg ttgacacaa 1080
tagatgctgt ccgcctgttg atgaggaagg gagcagaccc aagtactcgg aacttgga 1140
acgaacagcc agtgcatatt gttcccgatg gccctgtggg agaacagatc cgacgtatcc 1200
tgaagggaat gtccattcag cagagagctc caccgtatta gctccattag cttggagcct 1260
ggctagcaac actcactgtc agttaggcag tctgtatgta tctgtacata gaccatttgc 1320
cttatattgg caaatctaag ttgtttctat gacacaaaac tatttagttc actattatat 1380
acagt

```

<210> 54

<211> 1500

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2637522CB1

<400> 54

```

cacagagtga acaagagaga gtcatttggg aaacaaaagg agaattttac agagagagag 60
ggatagctaa aactacgtga gcctggcgag ggtgcagagc agaaagtaga gactgtccga 120
agactgctat ctgggacgag acaagttggt aaagggacag gagagaaaagc agagctattt 180

```

```

caagagtggag ccacagaagg gaatccagag gccatctaag cgaggaaggg tctacaggca 240
gtgagtgaag gccaggagca gggcccaggc caggcacgac caccgagggg atgaacttca 300
cagtgggttt caagccgctg ctaggggatg cacacagcat ggacaacctg gagaagcagc 360
tcattctgccc catctgcctg gagatgttct ccaaaccagt ggtgatcctg cctgccaac 420
acaacctgtg ccgcaaatgt gccaacgacg tcttccaggc ctggaatcct ctatggcagt 480
cccggggctc caccactgtg tcttcaggag gccgtttccg ctgcccacatg tgcaggcatg 540
aggttgtcct ggacagacac ggtgtctacg gcctgcagcg aaacctgcta gtggagaaca 600
ttatcgacat ttacaagcag gagtcatcca ggccgctgca ctccaaggct gagcagcacc 660
tcattgtgca ggagcatgaa gaagagaaga tcaatatatta ctgcctgagc tgtgagggtg 720
ccacctgtct tctctgcaag gtcttcggtg cccacaagga ctgtgagggt gccccactgc 780
ccaccattta caaacgccag aagagtggag tcagcgatgg catcgcgatg ctggtggcag 840
gcaatgaccg cgtgcaagca gtgatcacac agatggaggga ggtgtgccag actatcgagg 900
acaatagccg gaggcagaag cagttgttaa accagaggtt tgagagcctg tgcgcagtgc 960
tggaggagcg caacggtgag ctgctgcagg cgctggcccg ggaacaagca ggacaagctt 1020
caacgcgact cgacggact cattccggtc agtaattggg gaagcaactt ggggaaggccc 1080
tccttgccca aaagcctggt ttggagggtc actgtccaaa tccaagttt ccacatttgg 1140
atataccaga gcccaacaga aaattgtgtg cggacactta gtaaaatgcc cgtccaagag 1200
gccaagggtg cccaccaatg ggggagaacc tttgtacact ccacaggttt aaaaaagtgt 1260
tccccctga ggggccccac tatttttttc ccacataagc gatttgtggg aaaaaaactt 1320
tttgtgtcca cagagggggc cctcgggacc cccctgggga aaagtccac agaggcgctc 1380
ttaattttgt aatagaaggg gtccaaattt gtgcgaaaag ctccaaaaag tttcccaaac 1440
accgggggtt tataataccg aggggctttc gggggaaaaa gggccacccc cccttttttg 1500

```

<210> 55

<211> 2051

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2650980CB1

<400> 55

```

gcggacgccc aggtctcgct. tccgcccgtg cttccgtcgc cggggcgggc gccgcggtgt 60
ccctgccgag cgtgaccccc aactgtcacc gcgccgagcc ccagcctcat ccgtcggtgt 120
ccgcggttga ttcttcacca cactgaaacc attaggaaaa atccttgtgg ttaacagcag 180
aggcttcaga gtgtaacctg tactcggggc tagaaattat taaaatggc gactgatagc 240
tctcaagggt aactcgtcca tcctaaggca ctcccactta tagtaggagc tcagctgatc 300
cacgcggaca agttagggtga gaaggtagaa gatagcacca tgccgattcg tcgaactgtg 360
aattctaccc gggaaactcc tcccaaaagc aagcttgctg aaggggagga agaaaagcca 420
gaaccagaca taagttcaga ggaatctgtc tccactgtag aagaacaaga gaatgaaact 480
ccacctgcta cttcgagtga ggcagagcag ccaaaggggg aacctgagaa tgaagagaag 540
gaagaaaata agtccttctga ggaacccaaa aaggatgaga aagatcagtc taaagaaaag 600
gagaagaaag tgaaaaaaac aattccttcc tgggctaccc tttctgccag ccagctagcc 660
agggcccaga aacaaacacc gatggcttct tccccacgtc ccaagatgga tgcaatctta 720
actgaggcca ttaaggcatg cttccagaag agtgggtgat cagtgggttg cattcgaaaa 780
tacatcatcc ataatgatcc ttctctggag ctgagagaaa ggggttatct ccttaaaaac 840
gcaactgaaaa gagaattaaa tagaggagtc atcaaacagg ttaaaggaaa aggtgcttct 900
ggaagttttg ttgtggttca gaaatcaaga aaaacacctc agaaatccag aaacagaaag 960
aataggagct ctgcagtgga tccagaacca caagtaaaat tggaggatgt cctcccactg 1020
gcctttactc gcctttgtga acctaaagaa gcttcctaca gtctcatcag gaaatatgtg 1080
tctcagtatt atcctaagct tagagtggac atcaggcctc agctgttgaa gaacgctctg 1140
cagagagcag tagagagggg ccagttagaa cagataactg gcaaagggtg ttcgggggaca 1200
ttccagctga agaaatcagg ggagaaaccc ctgcttggtg gaagcctgat ggaatatgca 1260
atcttgtctg ccattgctgc catgaatgag ccgaagacct gctctaccac tgctctgaag 1320
aagtatgtcc tagaataatc cccaggaacc aattctaact atcaaagtc tttgctgaaa 1380
aaaaccctgc agaaaatgcga aaagaatggg tggatgggaa agatctctgg gaaagggttc 1440
agtggcacct tccagctctg ttttccctat tatcccagcc caggagtctt gtttccgaag 1500
aaagagccag atgattctag agatgaggat gaagatgaag atgagtcac agaagaagac 1560
tctgaggatg aagagccgcc acctaaagaa aggttgcaga agaaaacccc agccaagtcc 1620
ccaggaagg ccgcattctg gaagcagaga gggtcctaac ctgcacctaa agtctcagct 1680
gcccagcggg ggaaagctag gcccttgccct aagaaagcac ctcttaaggc caaaacgcct 1740
gccaagaaga ccagaccctc atccacagtc atcaagaaac ctagtgggtg ctctcaaag 1800
aagcctgcaa ccagtgcag aaaggaagta aaattgccgg gcaagggcaa atccaccatg 1860
aagaagtctt tcagagtga aaagtaaatt ttataggaaa aaagggtatc atgatgaaat 1920
tcaaaatctt attttctaag gtcagtgtgc atttgtttag ttttgatgct tttaaaatta 1980
cattattttc ctcccctatg aacattgtgg ggagggactc taaataaacc agtttaggca 2040

```

aaaaaaaaa a

2051

<210> 56
 <211> 1694
 <212> DNA
 <213> Homo sapiens

 <220>
 <221> unsure
 <222> 239
 <223> a or g or c or t, unknown, or other

 <220>
 <221> misc_feature
 <223> Incyte ID No: 2939607CB1

<400> 56
 gaattttgtg tttgcatcgt catcatgcgc gtgttgagcc ctttagtaat gtttttgtgt 60
 attagcttct tatttctagg tagacggtaa tttcgtcaat gaaaaaattt aacggtgac 120
 gtcacccctt gcctcagtc ctttctgtcc ctgcgcggc cccgggggtt accgcggacg 180
 ttctgttct gtggagaata acggctgcac ctgtatatt acgcatatt ggcggcgng 240
 gtgctttgtc tctcagcac tctgtgtca ctcaaggaag tatcatcaag aacaaggagg 300
 gcatggatgc taagtcacta actgcctggt cccggacact ggtgacctc aaggatgtat 360
 ttgtggactt caccagggag gagtggagc tgctggacac tgctcagcag atcgtgtaca 420
 gaaatgtgat gctggagaac tataagaacc tggtttctt gggttatcag cttactaagc 480
 cagatgtgat cctccggtt gagaaggag aagagccctg gctgggtggag agagaaattc 540
 accaagagac ccattcctgat tcagagactg catttgaaat caaatcatca gtttccagca 600
 ggagcatttt taaagataag caatcctgtg acattaaaat ggaaggaatg gcaaggaatg 660
 atctctggta tttgtcatta gaagaagtct ggaaatgtag agaccagtta gacaagtatc 720
 aggaaaaccc agagagacat ttgaggcaag tggcattcac ccaaaagaaa gtacttactc 780
 aggagagagt ctctgaaagt ggtaaatatg ggggaaactg tcttcttct gctcagctag 840
 tactgagaga gtatttccat aaacgtgact cacatactaa aagtttaaaa catgatttag 900
 ttcttaattg tcacaggac agttgtgcaa gtaacagtaa tgaatgtggt caaactttct 960
 gtcaaacat tcaccttatt cagtttgcaa gaactcacac aggtgataaa tctacaaat 1020
 gccctgataa tgacaactct cttactcatg gtcatctct tggatatca aagggcatac 1080
 atagagagaa accctatgaa tgtaaggaat gtggaaaatt cttcagctgg cgctctaata 1140
 ttactaggca tcagcttatt catactggag aaaaacccta tgagtgtaaa gaatgtggaa 1200
 agtcttacag ccagagatct cacttgttg tgcattcatg aattcacact ggactaaaac 1260
 cttttgagt taaggattgt ggaaaatgtt ttagtcgaag ctctcacctt tattcacatc 1320
 aaagaaccca cactggagag aaaccatatt agtgtcatga ttgtggaaaa tctttcagcc 1380
 agagtcttgc ccttattgtg catcagagga tacacactgg agagaaacca tatgaatgct 1440
 gtcagtgtgg gaaagcctt atccgggaaga atgacctcat taagcaccag agaattcatt 1500
 ttggagaaga gacctataaa tgtaatcaat tgggcattat cttcagccag aactctccat 1560
 ttatagttca tcaaatagtc cacactggag agcagttctt aacatgcaat caatgtggga 1620
 cagcgttgt taatacctct aaccttattg gataccagac aaatcatatt agagaaaaag 1680
 cttactaata aata 1694

<210> 57
 <211> 1177
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3098421CB1

<400> 57
 gacctggccc ccgaatcagg aaaaggcagg tattggaacc cacggcctgg gtgttagtcc 60
 agccgcttag tattctgact ctattctgcc tttctgtct cctaagaata actgtgcttg 120
 aagaagaaaa ttccaacat ggacaaacca cgcaagaaa atgaagaaga gccgcagagc 180
 gcgccaaga ccgatagga gaggcctccg gtggagcact ctcccgaata gcagtccttc 240
 gaggagcagt cttcggagga gactgctcgt gaggaggagt tctttcctga gaggctcttg 300
 cctgagctcc tgcctgagat gctcctctcg gaggagcgc ctccgcagga gggctttcc 360
 aggaaggacc tgtttgagg gcgcctccc atggagcagc ctcttgttg agtaggaaaa 420
 cataagcttg aagaaggaa ctttaagaa aggttggctc gttctcgcgc gcaatttaga 480
 ggggacatac atggcagaaa ttaagcaat gaggagatga tacaggcagc agatgagcta 540
 gaagagatga aaagagtaag aaacaaactg atgataatgc actggaaggc aaaacggagc 600

```

cgctccttate ctattttaatg tgttcggcct ttaattctgt tttgcctgct aatagtattg 660
ccattggccac ctggactttc tgtttgcat tctttaatgc cttttcccat attctgaatt 720
ttaacttttt gtgaggcttt atttttagat ttttagcatg aactcgctta aagttgaggt 780
ttccccctaa aatctacaag ttccctctt tcagtcata gacctacaca tttgcatgaa 840
agatgtacat tatataattgt gaaacgaaaa aagcaatttt caaatggtat atattgtatc 900
ccatttttgt aaaaaaatg tatatttata tattaatatg caaagaaaaa gctaaaaagta 960
tagacttcaa aggcataaca gtggttgtgt ggtaagataa taggtgattt tttaaatttt 1020
tgttttatct gaattttctca ttttttcagg acaaacgttt tacttggtgt gcaaaaaatat 1080
ataatgaaaa aatcacacaa ttttgaagaa aactgtcaat cagcttataa cgacaatgtg 1140
gcacttaata aatacttgtc agaactttaa aaaaaaaa 1177

```

<210> 58

<211> 1219

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> 1139

<223> a or g or c or t, unknown, or other

<220>

<221> misc_feature

<223> Incyte ID No: 3296650CB1

<400> 58

```

gctcggaatt cggctcgagg tagccagtgc cagccaatga aggctctgtt caagcatgaa 60
tctctgggat cccagccctt acacgataga gttctccagg ttcctgggct tgcccagggg 120
gggtgctgca gagaagatgc aatggtagct tccagggtca ctccagggtc ccagggtttg 180
ctgaaaatgg aagatgtggc cctgacctc actcctgggt ggacacagct ggattcatct 240
caggtgaacc tctacagaga tgaaaagcag gagaaccata gcagcctggt ctcccttggt 300
ggtgaaatac agactaagag cagggacttg cctccagtca agaagcttcc agaaaaggag 360
catgggaaga tatgccacct gaggggaagac attgccaga ttcctacaca tgcagaagct 420
ggtgaacagg agggcaggtt acaaagaaaag cagaaaaatg ccatagggag taggcgacat 480
tattgccatg aatgtggaaa gagttttgct caaagttcag gcctgactaa acacaggaga 540
atccacactg gtgagaaacc ctatgaatgt gaagactgtg gaaagacctt cattgggagc 600
tctgccttgg tcattcatca gagagtccac actggtgaga agccatatga gtgtgaagaa 660
tgtggttaagg tcttcagtca cagctcaaac cttatcaaac accagagaac ccacactggg 720
gagaagccct atgagtgtga tgactgtggg aagaccttca gccagagctg cagcctcctt 780
gaacatcaca aaattcatac tgggggagaag ccataccagt gcaatatgtg tggcaaagcc 840
tttaggcgga attcacatct cctgagacat cagaggattc atggtgataa aaatgttcag 900
aatcctgagc acggggagtc ctgggaaagt cagggtagga cggaaaagcca tggggaaaat 960
actgaggctc ccgtgtctta taaatgtaat gagtgtgaga gaagtttcac acggaataga 1020
agtcttattg aacatcagaa aatccacact ggtgacaaac cctatcagtg tgacacatgt 1080
ggaaaagggt tcaactgaac ttcatacctt gttcaacatc agagaagcca tgtagggana 1140
aaaactcttt cacagtgacc catggttatc atgccaacat tgggtgctcat ttgtcactga 1200
actgaagcaa cccttgagg 1219

```

<210> 59

<211> 1309

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3687719CB1

<400> 59

```

ctggaagctg agcctgaccc agcagggtccc agcagcctgg cagcccggcc agcatgcagc 60
agcagcctct gcccgggcct ggcgccccca caactgagcc aaccaagcct ccctacagct 120
acatgcacct tattgctatg gccatccaga gctcaccggg gcagcggggc accctcagtg 180
gcatctaccg ctacatcatg ggccgattcg ccttctaccg ccacaaccgg cccggctggc 240
agaacagcat ccgccacaat ctgtcactca acgagtgtt tgtcaagggt ccccgcgatg 300
accgcaagcc aggcaagggc agctactgga cgctggaccc tgactgccac gacatgtttg 360
agcacggcag cttcctacgc cgccgcccgc gcttcacccg gcagacaggt gctgagggca 420
cccgggggcc cgccaaggca cgccgtggac cctcaggggc gaccagccag gaccagggag 480
tccccaacgc cagcaccggc aggcagtgtc cattcccacc agagctgcca gatcccaagg 540

```

```

gcctaagctt  tgggggtctg  gtggggggcca  tgccagccag  tatgtgcccc  gcaaccactg  600
atggcaggcc  tgcggccaccc  atggagcccca  aagagatttc  cacgcccacg  cctgcatgcc  660
caggggagct  ccccggtggc  acctcatctt  cctcatgccc  agcgtttggc  tttcctgccc  720
gcttctcaga  ggctgagagt  tttaataaag  cccctacgcc  cgtcttgccc  ccggaatcag  780
gcatcgggag  cagctaccag  tgtcggctgc  aggcactgaa  tttttgcatg  ggggctgacc  840
caggccttga  gcacctcttg  gcctcagcag  cccctcccc  tgcaccaccc  acccctccag  900
gctcactccg  ggccccactg  cccctgccaa  ctgaccacaa  ggaaccctgg  gttgcagggtg  960
gcttccctgt  ccaggggagg  tccggctacc  cattgggggt  gacccctgc  ctataccgga  1020
cgccaggaat  gttcttcttt  gagtaaaagg  agcctcacct  cgggcagtc  ctgcagggtcc  1080
ctcacctcc  gggctgagcc  tggctctagg  acctgaagaa  ctctcgaagg  accagccct  1140
ggtgagccaa  ggacaaccac  acagaaagcc  aggattgaag  cggtgctcag  ccaggccct  1200
ggggcctccg  gacaaacttg  ggggtgaggg  gaagcagggc  ctcttgggat  ttactctgtg  1260
gctctcaggg  ccaataaagc  catgtgatga  tgaaaaaaa  aaaaaaagg  1309

```

<210> 60
 <211> 1326
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 3774188CB1

```

<400> 60
gaggagctca  actgatctgt  tttcttttcg  ccagccaaaa  tcacagaatg  aaggcgggtga  60
agagcgaacg  ggagcagagg  agccggcgaa  gacaccggga  cggggagcgt  gtgctgccgg  120
cgggggtggt  ggtgaagcag  gagcgtctca  gcccagaagt  cgacacctcc  gccaccgcc  180
gtccggacca  ctccggtggt  agcccgcttc  cgccgaccag  cgagccggcc  cgctcgggcc  240
accgcgga  ccgagcccg  ggagttagcc  ggtccccacc  caaaaagaaa  aacaaggcct  300
cagggagaag  aagcaagtct  cctcgcagta  agagaaaccg  aagtcctcac  cactcaacag  360
tcaaagtga  gcaggagcgt  gaggatcatc  cccggagagg  acgggaggat  cggcagcaca  420
gggaaccatc  agaacaggaa  cacaggagag  ctagggaacag  tgaccgggac  agacaccggg  480
gccattccca  ccaaaggaga  acgtctaacc  agaggcctgg  gaggggcag  ggtcagggac  540
gggatcgaga  cactcagaac  ctgcaggctc  aggaagaaga  gcgggagttt  tataatgcca  600
ggcgacggga  gcatcgccag  aggaatgacg  ttggtggtgg  cggcagttag  tctcaggagt  660
tggttcctcg  gcctggtggc  aacaacaaag  aaaaagaggt  gcccgcataa  gaaaaaccaa  720
gctttgaact  ttctggggga  cttcttgagg  acaccaacac  ttccgggggt  gtatgcatta  780
aatatagtga  gcccccagaa  gcacgtatcc  ccaaaaaaac  gtggcgcttc  taccattta  840
aaaatgatga  ggtgcttcca  gtcagtaca  tacatcgaca  gagtgcgtac  ctactgggtc  900
gacaccgccg  cattgcagac  attccaattg  atcacccgtc  ttgttcaaag  cagcatgcgg  960
tctttcaata  tcggcttgtg  gaatatacc  gtgctgatgg  cacagtggc  cgaagagtga  1020
agcccatcat  cattgacctt  ggctcaggca  attggaacct  cttaaacaac  aaacgtattg  1080
agccacagag  atactatgaa  ctaaaagaaa  aggatgtact  caaatttga  ttcagtagca  1140
gagaatacgt  cttgctccat  gagtgcgcgg  acacttctga  aatagacagg  aaagatgacg  1200
aggatgagga  ggaggaggaa  gaagtgtctg  acagctagca  aactaagaac  ccaaactatt  1260
gatacacggt  ttcttctctg  gaagtctttg  attgactcag  agagcactat  ggtggtgggt  1320
ccagca  1326

```

<210> 61
 <211> 1097
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 4349106CB1

```

<400> 61
gcggcttccg  ggatttggcg  gtggcctttg  ttggctgcag  taagagctca  gtctcttcac  60
caggggtccc  cagtccttcc  atctgggagg  ccaaggcggc  ttcgcgttct  gagaatagac  120
agaacctctg  ttactctgtg  accggcaggc  accgggagat  ccgtagctca  gacgccagga  180
catcccgga  cgtgggaaat  gggactgttg  acattcaggg  atgtggccat  agaattctct  240
cgggaggagt  gggaaacacct  ggactcagat  cagaagcttt  tatatgggga  tgtgatgtta  300
gagaactacg  gaaacctggg  ctctctgggt  ctgcgtgtct  ctaagccgga  cctgatcacc  360
tttttgagc  aaaggaaaga  gccctggaat  gtgaagagt  cagagacagt  agccatccag  420
ccagatatct  ttctcatga  tactcaaggc  ctcttaagaa  agaagcttat  agaagcatca  480
ttccaaaaag  tgatatttga  tggatatggg  agctgtggcc  ctcagaattt  aaacttaagg  540

```

```

aaagagtggg aaagtgaggg caaaataatc ctatggtgaa aaaaaatcaa caagataatc 600
tctgcatgag aaaaaggacc aaagacaagt gaattttctg agggatgatga aagtgttgct 660
tcgaaaagggt gtgaagttta tatgggtcta tttattttgtc taacatgtac agttaagggt 720
tatgccttgc aatgtatgta catcttcaca aaaaaaatct taaaaaaatt aaatgggtgg 780
gggtagggaa agggttgaag tatagatgaa gcagaagtgg tacatgatta gtagtgaag 840
ctgggggcag gtctatatat tttattttta tgtctttaat agcatttgta taaatgtaca 900
atattcgttt acaatgttag ctcaggatct tgtttacct tggacaggga gggagggaga 960
aatttatttt ctgggtagga gtaaagctgg ttctattcct cagttatgta aattatctgt 1020
tataatccaa aaagctttac atgaatgttt ctgtatgtaa tgtggtatat caataaaaaa 1080
ttaaatatta aaaaaaa 1097

```

<210> 62

<211> 2404

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 4834217CB1

<400> 62

```

ctcgcgtggct agtaggagag actggtgctt gccccgcccc gtggactaac tcgcttaatt 60
ttaaataaaaa agtcgaggac acggcggtcg ttttcccga gacatggggc ctcccatggg 120
ccatttgcct cctggaggcc ctcgcgtctt gctgagcccc gggagttagg atgacgcgag 180
cgggtgagga gcccggaacg attccttcgc ggaacaattg aggcgaggcc tttgggagta 240
ctttgtggga cggaccctgg cgggcccctgc cagacgcaca gggatggcgg cggaggcggc 300
cgatttgggg ctggggggcc cgcgtcccctg ggagctgagg cgggagcgac gcatggtgtg 360
cgtggagtag ccgggagtg tgcgtgatgt ggctaagatg ctgccgactc tgggcggcga 420
ggaaggcgct tcccggatct acgcagacct caccaagagg ctggagctgt acttccggcc 480
caaggaccca tactgccacc cagtgtgcgc caaccgcttc agtaccagca gcctgctgct 540
ccgcatcagg aagagaacga ggcggcagaa aggggtgctg ggcactgagg cccactccga 600
ggtcacattt gacatggaga tccttggcat catctccacc atttacaat ttcaggggat 660
gtctgacttc cagtacttgg ctgtgcatac ggaagcagga ggcaagcata cgtcaatgta 720
tgacaagggt ctcatgctcc ggcccagaaa ggaggccttt ttccaccagg agctgccgct 780
ctacatcccc ccaccatct tctcccggct ggacgccccg gtggactact tctaccgacc 840
agagacccag caccgggaag gctacaacaa tccccccatc tcaggtgaga atctgatttg 900
cctgagcaga gcccgggccc cccacaatgc catctttgtc aactttgagg atgaggaggt 960
gcccagcag ccactggagg ctgcagccca gacgtggagg agagtctgca ctaaccctgt 1020
ggaccgggaag gtggaggagg agctgaggaa gctgtttgac atccgtccca tctggtcccc 1080
aaatgctgtc aaggccaaca tcagcgtcca ccagacaag ctcaaggctt tgcttccctt 1140
catagcctat tacatgataa caggcccctg gcgcagccta tggattcgat ttgggtatga 1200
ccccgaaaa aaccagatg ccaagattta tcaagtcctc gatttccgaa tccgttgtgg 1260
aatgaaacac ggttacgccc ccagtgaact gccggtcaaa gcaaagcgca gcaacctaaa 1320
ctacagcctc cccatcaccc tcaagaagac atccagccag cttgtcacca tgcattgact 1380
gaagcagggc ctggggccgt cggggacgag tgggtgctcg aaaccagctt ccagcaagta 1440
caagctcaag gactctgtct acatcttccg ggaagggggc ttgccaccct atcggcagat 1500
gttctaccag ttatgcgact tgaatgtgga agagtgcag aagatcattc accgcaatga 1560
cggggcagag aattcctgca cagaacggga tgggtggtgc ctccccaa ga ccagcgacga 1620
gctcagggac accatgtccc tcatgatccg gcagaccatc cgctccaaga ggcctgctct 1680
cttttccagc tcagccaagg ctgatggcgg aaaagagcag ctgacgtaag agtctgggga 1740
agacgaggag gatgaggagg aggaggaaga ggaggaggag gacttcaagc catccgacgg 1800
cagtgaaaaa gaaatggaga cagagattct ggactacgtg tgacagggcc caaggctggg 1860
cctccctgac ccggccagac tgggtgtctg cctaattagg gagccggggc tccccattgc 1920
caccacagat gcccggaatg gccctaggag gccctctgag gagagctaga gtcccagcaa 1980
aggggtgcagc tgaccctagc actggctgtg acatgctgct tgggtgctgc tctggtcctg 2040
aggggttagg gacatcccca aagggtatac cctggctctg ccacccatga accagcccag 2100
catccagcca gtgagtgggc acccaatgcc tctcaggatg agaccagtaa atgccggagg 2160
tggagctggg cagctgtgga gcccaggcc acaggccagt ctgcttggc tctcatgact 2220
gtggtggtgg agatagcgtg gggagcctcg cccatggtct cacgtggcaa gaagtgcctt 2280
tagctctgga tcccaaccgt ttggcacagc tttggccaca gccaggcccc tctggaattg 2340
tccttattaa accagtttcc cgagaaaaaa aagaataaaa acagcagaaa aaataaaaaa 2400
aaaa 2404

```

<210> 63

<211> 1900

<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5156094CB1

<400> 63

agctggagca	ccgtgaggaa	gaagcgaggt	tctttttaag	agttcagctg	cgaggctgta	60
acagaggagg	aaatgatttt	gaatagcctc	tctctgtgtt	accataataa	gctaatacctg	120
gccccaatgg	ttcgggtagg	gactcttcca	atgaggctgc	tggccctgga	ttatggagcg	180
gacattgttt	actgtgagga	gctgacgcac	ctcaagatga	ttcagtgcaa	gagagttgtt	240
aatgaggtgc	tcagcacagt	ggactttgtc	gccccctgatg	atcgagttgt	cttccgcacc	300
tgtgaaagag	agcagaacag	ggtggtcttc	cagatgggga	cttcagacgc	agagcgagcc	360
cttgctgtgg	ccaggcttgt	agaaaatgat	gtggctggta	ttgatgtcaa	catgggctgt	420
ccaaaacaat	attccaccaa	gggaggaatg	ggagctgccc	tgctgtcaga	ccctgacaag	480
attgagaaga	tcctcagcac	tcttgttaaa	gggacacgca	gacctgtgac	ctgcaagatt	540
cgcatacctgc	catcgctaga	agataccctg	agccttgtga	agcggataga	gaggactggc	600
attgctgcca	tcgcagttca	tgggaggaag	cgggaggagc	gacctcagca	tcctgtcagc	660
tgtgaagtca	tcaaagccat	tgctgatacc	ctctccattc	ctgtcatagc	caacggagga	720
tctcatgacc	acatccaaca	gtattcggac	atagaggact	ttcgacaagc	cacggcagcc	780
tcttccgtga	tgggtggccc	agcagccatg	tggaaacccat	ctatcttctt	caaggagggt	840
ctgcggcccc	tggaggaggt	catgcagaaa	tacatcagat	acgcggtgca	gtatgacaac	900
cactacacca	acaccaagta	ctgcttgtgc	cagatgctac	gagaacagct	ggagtgcgcc	960
cagggaaaggt	tgctccatgc	tgcccagttc	tcccgggaaa	tttgtgaggc	ctttggcctt	1020
ggtgccttct	atgaggagac	cacacaggag	ctggatgccc	agcaggccag	gctctcagcc	1080
aagacttcag	agcagacagg	ggagccagct	gaagatacct	ctggtgtcat	taagatggct	1140
gtcaagtttg	accggagagc	ataccagccc	cagatcaccc	ctaagatgtg	cctactagag	1200
tgggtgccgga	gggagaagtt	ggcacagcct	gtgtatgaaa	cggttcaacg	ccctctagat	1260
cgcctgttct	cctctattgt	caccgttgct	gaacaaaagt	atcagtctac	cttgtgggac	1320
aagtccaaga	aactggcgga	gcaggctgca	gccatcgtct	gtctgcggag	ccagggcctc	1380
cctgagggtc	ggctgggtga	ggagagccct	tccttgacaca	agcgaagag	ggaggctcct	1440
gaccaagacc	ctggggggccc	cagagctcag	gagctagcac	aacctgggga	tctgtgcaag	1500
aagccctttg	tggccttggg	aagtggtgaa	gaaagccccc	tggaaaggctg	gtgactactc	1560
ttcctgcctt	agtcacccct	ccatgggcct	ggtgctaagg	tggctgtgga	tgccacagca	1620
tgaaccagat	gocgttgaac	agtttgctgg	tcttgccctgg	cagaagttag	atgtcctggc	1680
aggggccatc	agcctagagc	atggaccagg	ggccgcccag	gggtggatcc	tggccccttt	1740
ggtggatctg	agtgacaggg	tcaagttctc	tttgaaaaca	ggagcttttc	aggtggtaac	1800
tccccaacct	gacattggta	ctgtgcaata	aagacacccc	ctaccctcac	ccacggctgg	1860
ctgcttcage	cttggggcatc	ttcataaaaa	aaaaaaaaaa			1900

<210> 64
<211> 2901
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5665139CB1

<400> 64

accacagcgc	agtctcactc	tgtccttcag	gctgtgagtg	cagtctcatg	atcttggctc	60
actgcaacct	ctgtctcctg	ggttcaagtg	attcttgtgc	ctcagcctac	caagtagctg	120
agatacagga	ttgacttcta	aagactcttg	gtacctgagg	aagaaaccca	gaagagggaag	180
aggaaagcaa	aggagtcggg	gatggctctt	tctcagggtc	tattgacatt	cagggacgtg	240
gccatagaat	tctctcagga	ggagtggaaa	tgccctggacc	ctgctcagag	gactctatac	300
agagacgtga	tgctggagaa	ttataggaac	ctggagtctg	tggggtctat	tgacattcag	360
ggatgtggcc	atagaattct	ctcaggagga	gtggaaatgc	ctggaccctg	ctcagaggac	420
tctatacaga	gatgtgatgt	tggagaatta	taggaacctg	gtctccctgg	gtattttttc	480
taaatgtgag	atcaaggaat	taccaccaaa	aaaggagagt	aatacaggag	aaatattcca	540
gacagtaatg	ttggaaagac	atgaaagcca	cgacatacaa	gatttttgct	tcagagaaac	600
ccagaaaaat	gtacatgact	ctcagtgtct	gtggaaacat	gattgaagac	attataagcg	660
agtgcggtgtg	acctataagg	aaagtctcat	tggtagaaga	gacatgcatt	gtagaaagga	720
tgatgcacaa	aagcagcctg	ttaaaaatca	gcttggatta	aacccgcagt	cacatctacc	780
agaactgcag	ctatttcaag	ctgaaggga	aatatataaa	tatgatcaca	tggaaaaaac	840
tgtcaacagt	agttccttag	tttccccacc	ccaacgtatt	tcttctactg	tcaaaaccca	900
catttctcat	acatatgaat	gtaattttgt	ggattcatta	ttcacacaaa	aagagaaagc	960
aaatattggg	acagaacact	acaaatgtaa	tgagcgtggc	aaggcctttc	atcaaggctt	1020
acattttact	atacatcaaa	taatccatac	taaagagacg	caattttaa	gtgatatatg	1080
tggcaagatc	ttcaataaaa	aatcaaacct	tgcaagtcac	caaagaattc	atactggaga	1140

gaagccatat	aaatgtaatg	aatgtggcaa	ggctcttccat	aatatgtcac	accttgcaca	1200
gcatcgcagg	attcatactg	gagagaaacc	atataaatgt	aatgaatgtg	gcaaggtctt	1260
taatcaaatt	tcacaccttg	cacaacatca	aagaattcat	accggagaga	aaccttataa	1320
atgtaatgaa	tgtggaaagg	tcttccatca	aatttcacac	cttgcacaac	atcggacaat	1380
tcatactgga	gaaaaacctt	acgaatgtaa	caaattgtggc	aaggtgttca	gtcgcgaattc	1440
ctaccttgta	caacatctga	tcattcatac	tggagagaaa	ccttacagat	gtaatgtatg	1500
tggaaagggtc	ttccatcata	tttcacacct	tgcacaacat	cagagaatcc	acactggaga	1560
gaaaccttac	aaatgtaatg	agtgtggcaa	ggctcttcagt	cacaagtcac	ccctagttaa	1620
tcactggaga	attcatactg	gagagaaacc	ttacaaatgt	aatgagtgtg	gcaaggtctt	1680
cagtcacaag	tcattccctag	taaatcactg	gagaatccac	actggagaga	aaccttacia	1740
atgtaatgaa	tgtggcaagg	tcttcagtcg	caattcatac	cttgcccaac	atctgataat	1800
tcattgccggt	gagaaacctt	ataagtgtga	tgaatgtgac	aaagcattca	gtcaaaaattc	1860
acattcttgta	caacatcaca	gaatccatac	tggagagaaa	ccttaciaaat	gtgatgaatg	1920
tggcaaaagtc	ttcagtcaaa	attcatacct	tgcataatcat	tggagaattc	atactggaga	1980
aaaagcttat	aaatgtaatg	aatgtgggaa	ggctcttcggt	ctaaactcat	ccctagcaca	2040
tcattcgga	attcacactg	gagagaaacc	tttcaaatgt	aatgaatgtg	gcaaagcttt	2100
tagtatgcgt	tcaagcctca	ctaattcatca	tgcgatccac	actggagaga	aacatttcaa	2160
atgtaatgaa	tgtggcaaac	tcttccgcga	caattcatal	cttgtacgtc	atcagagatt	2220
tcattgccgga	aagaaatcta	acacatgtaa	ttaatgtggc	agagtgttca	gttagcatta	2280
aagccttgta	agacatacaa	taattttatac	tggagaaaaa	ctttgcaagt	ataatgaatg	2340
tagcagagcc	tttagttttt	gttcaaggct	taataaccgt	tagctagacc	atagaggaca	2400
gaaactttac	taatgtactg	aatgtggcaa	ggctcttaagg	taaaatctga	gaccaggatt	2460
tttcaaagaa	ttcttgctgg	tgagaaacct	aacaaatgta	atgaatgtgg	caaggtcttc	2520
tggcacaatt	ctcacattgt	acaatattgc	aaaaattcat	gcttgagaga	aacaaaaaca	2580
ctgagagtgg	gaaaccatta	tgacttcaaa	cattcatcaa	catcagagaa	tcataactaa	2640
agagcatttta	taataattat	atgtgataga	gattttccgc	aggccaaagt	ctcactaggc	2700
atcaaaaact	tttttgatga	aaccatacaa	atgtaacgtg	catgcttaag	cttttacccta	2760
ggcatcaaaa	ccggaacatc	acaggggttta	tactggagag	taactacaca	aagataatgt	2820
aataagcctt	tcagtgtaat	attcatgatt	ttgtcgtgag	agatccactc	aataaaaaacc	2880
aggcaaatgt	aaaaaaaaaa	a				2901

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 December 2000 (28.12.2000)

PCT

(10) International Publication Number
WO 00/78954 A3

(51) International Patent Classification⁷: C12N 15/12,
C07K 14/47, 16/18; C12Q 1/68, G01N 33/68, A01K
67/027, A61K 38/17

(US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs
Drive, Hayward, CA 94545 (US). TRAN, Bao [US/US];
750 Salberg Avenue, Santa Clara, CA 95051 (US).

(21) International Application Number: PCT/US00/16766

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics,
Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(22) International Filing Date: 15 June 2000 (15.06.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/140,109 18 June 1999 (18.06.1999) US

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE,
ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP,
KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD,
SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ,
VN, YU, ZA, ZW.

(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/140,109 (CIP)
Filed on 18 June 1999 (18.06.1999)

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): INCYTE
GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo
Alto, CA 94304 (US).

Published:
— with international search report

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): LAL, Preeti
[IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US).
YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA
94087 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick
Court, San Jose, CA 95118 (US). BAUGHN, Mariah, R.
[US/US]; 14244 Santiago Road, San Leandro, CA 94577

(88) Date of publication of the international search report:
23 August 2001

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

WO 00/78954 A3

(54) Title: HUMAN TRANSCRIPTIONAL REGULATOR PROTEINS

(57) Abstract: The invention provides human transcriptional regulator proteins (TXREG) and polynucleotides which identify and encode TXREG. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of TXREG.

INTERNATIONAL SEARCH REPORT

Intern al Application No

PCT/US 00/16766

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/47 C07K16/18 C12Q1/68 G01N33/68
A01K67/027 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K C12N C12Q G01N A01K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] EBI; ACC.NO.: AA023950, 14 August 1996 (1996-08-14) MARRA ET AL.: "The WashU-HHMI Mouse EST project" XP002152509 abstract	1-19,22, 25-27
X	DATABASE EMBL [Online] EBI; ACC.NO.: AI509143, 17 March 1999 (1999-03-17) MARRA ET AL.: "The WashU-NCI mouse EST project" XP002152510 abstract	1-19,22, 25-27

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

10 November 2000

Date of mailing of the international search report

27.02.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

van Klompenburg, W

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/16766

C.(Continuation), DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>RAZDAN ET AL.: "Molecular cloning of a novel platelet protein showing homology to the angiotensin II receptor C-terminal domain"</p> <p>THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 4, 26 January 1996 (1996-01-26), pages 2221-2224, XP002152508 figure 1</p>	5,11-15
X	<p>--- DATABASE EMBL [Online] EBI; ACC.NO.: AA158198, 19 January 1997 (1997-01-19) HILLIER ET AL.: "Generation and analysis of 280,000 human expressed sequence tags" XP002152511 abstract</p>	5,11-15
A	<p>--- US 5 861 495 A (COLEMAN ROGER ET AL) 19 January 1999 (1999-01-19) seq id nos:3,4 column 1, line 11 -column 3, line 11 claims 1-8; examples 1-12</p>	1-19,22, 25-27
A	<p>--- US 5 726 288 A (DARVEAU ANDRE ET AL) 10 March 1998 (1998-03-10) column 2, line 20 -column 3, line 18 column 6, line 14 - line 59 column 22, line 4 - line 40</p>	1-19,22, 25-27
A	<p>--- REUTER S ET AL: "APM-1, A NOVEL HUMAN GENE, IDENTIFIED BY ABERRANT CO-TRANSCRIPTION WITH PAPILLOMAVIRUS ONCOGENES IN A CERVICAL CARCINOMA CELL LINE, ENCODES A BTB/POZ-ZINC FINGER PROTEIN WITH GROWTH INHIBITORY ACTIVITY" EMBO JOURNAL,GB,OXFORD UNIVERSITY PRESS, SURREY, vol. 17, no. 1, 1 January 1998 (1998-01-01), pages 215-222, XP000867721 ISSN: 0261-4189 abstract; figure 3</p> <p>-----</p>	1-19,22, 25-27

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 00/16766

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
Although claim 18 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 20,21,23,24
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Claims 1-27, all partially.

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 20,21,23,24

Claims 20,21,23 and 24 refer to an agonist and antagonist of the polypeptide without giving a true technical characterization. Moreover, no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: Invention 1: Claims 1-27, all partially

Polypeptide according to seq.ID.1 or having 90% homology thereto, biologically active and immunogenic fragments thereof, polynucleotide encoding it (e.g. SEQ.ID.NO.33) or fragments thereof comprising at least 60 contiguous nucleotides, vector comprising said polynucleotide, host transformed with said vector and its use for producing said polypeptide, antibody against said polypeptide, method for detecting said polynucleotide through hybridization, method for identifying an (ant)agonist or a compound that alters the expression of said polypeptide, and a pharmaceutical composition of said polypeptide or said (ant)agonist.

2. Claims: Inventions 2 to 32: Claims 1-27, all partially

Subject matter analogous to that defined above, under invention 1, but limited to the respective proteins with SEQ ID NOs: 2-32 and the nucleic acids encoding them, represented by SEQ ID NOs: 34-64.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/16766

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5861495 A	19-01-1999	AU 6035298 A	07-08-1998
		WO 9831805 A	23-07-1998
US 5726288 A	10-03-1998	DE 69033127 D	01-07-1999
		DE 69033127 T	14-10-1999
		EP 0453560 A	30-10-1991
		JP 4503014 T	04-06-1992
		WO 9107509 A	30-05-1991
		US 5350840 A	27-09-1994

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.